

DOCUMENT RESUME

ED 233 037

TM 830 434

AUTHOR King, J. A.; Perin, E. M.
TITLE The Process of Evaluation Use in Local School Settings. Final Report.
INSTITUTION New Orleans Public Schools, LA. Dept. of Research and Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Oct 82
GRANT NIE-G-81-0900
NOTE 83p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Administrators; Decision Making; Elementary Secondary Education; *Evaluation Methods; *Evaluation Needs; *Evaluation Utilization; Evaluators; *School Districts; *Urban Areas; Use Studies
IDENTIFIERS Conceptual Frameworks; Naturalistic Evaluation

ABSTRACT

This report presents the results of a yearlong study of the process of evaluation use in a large city school district. Based on extensive naturalistic data-gathering, the research documents how local school administrators use evaluation information generated by the district's research and evaluation unit. The significance of the research is the framework presented for conceptualizing the evaluation use process in a local education agency, including the following concepts: three new or revised use factors ("espoused theories" vs. "theories-in-use," an expanded personal factor, and the clout factor); the types of use (signalling vs. charged use); the evaluation use process; and the domain of charged use. Concluding recommendations are directed toward local school evaluators and users. They emphasize the importance of evaluator sensitivity to organizational complexities within school districts, urging that most time and resources should be spent on evaluations supported by key administrators who have internal clout. Evaluation reports required to "signal" that the ongoing program responsibility is met will be conducted to fulfill external requirements, but these reports may not have the potential to directly effect local program change. Finally, implications for future research, including the call for more naturalistic evaluation use studies, are presented. (Author/PN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED233037

THE PROCESS OF EVALUATION USE IN LOCAL SCHOOL SETTINGS

Final Report of NIE Grant 81-0900

ELLEN PECHMAN, DIRECTOR
TESTING & EVALUATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

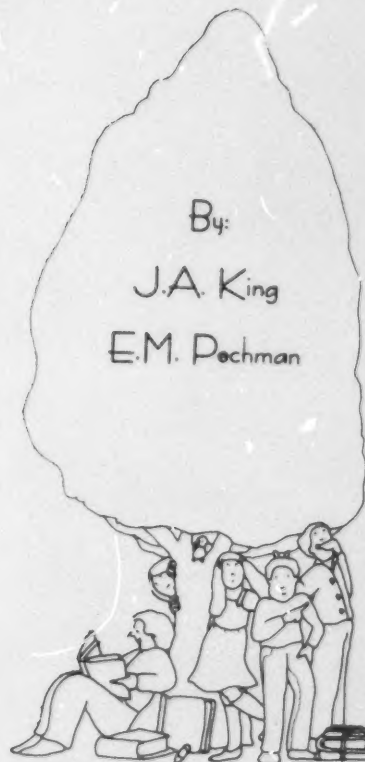
E. PECHMAN

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

NEW ORLEANS PUBLIC SCHOOLS
4100 TOURO STREET
NEW ORLEANS, LA 70122
(504) 288-6561

CHARLES E. MARTIN, SUPERINTENDENT

OCTOBER 1982



The Process of Evaluation Use
in Local School Settings

Final Report of NIE Grant NIE-G-81-0900

J. A. King

Tulane University

E. M. Pechman

Orleans Parish School Board

October, 1982

The research discussed here was supported by the National Institute of Education;
however, no endorsement by the Institute is implied.

The Process of Evaluation Use in
Local School Settings

Table of Contents

	Page
Executive Summary	i
Abstract	iii
List of Illustrations	iv
Acknowledgments	v
Foreword and Rationale	1
Section 1. Introduction	6
Section 2. The Research Setting and Methodology	9
Section 3. Rethinking Local Evaluation Use	17
Section 4. A Conceptual Framework for Evaluation Use	34
Types of Use	35
The Process of Use	47
The Domain of Charged Use	51
Section 5. Summary and Implications	60
Bibliography	67

The Process of Evaluation Use in Local Schools

(NIE-G-81-0080)

Executive Summary

This report presents the results of a yearlong study of the process of evaluation use in a large city school district. Based on extensive naturalistic data-gathering, the research documents how local school administrators use evaluation information generated by the district's research and evaluation unit. The report then presents a framework for conceptualizing the evaluation use process in an LEA.

The findings of the study question recent assumptions about evaluation reporting procedures and collaboration as remedies to the problem of use. The data suggest that these may work in some instances, but not in others because both the use and non-use of evaluation were found to be viable and "rational" alternative responses to the evaluation process in the complex political environment of a school bureaucracy.

To the factors contributing to evaluation use suggested in the recent literature, this study added or revised three concepts. First, we discovered a distinction between what users said they believed in and what they acted upon and apparently believed in. This difference between "espoused theories" and "theories-in-use" means that evaluation processes and products may be used by decision-makers, but not in a direct and predictable manner. Second, the personal factor discussed by Patton et al. (1977) and others was expanded to include dimensions of the self-confidence of the primary users and the receptivity of the organizational context. Third, we found that high level managers with decision authority, the people with clout, must care about the evaluation process and assure it moves forward if evaluation is to be used.

The report also describes a conceptual framework for discussing the evaluation use process. Two "types" of evaluation use are presented: "signalling," whereby information is sent out of the system to signal that required activities are proceeding as mandated; and "charged," whereby the user takes the information and uses it either instrumentally or persuasively. The process of evaluation use was found to include evaluation activities, evaluation use products, i.e., users' actions and their changes in attitude, and the written evaluation products -- informal reports (e.g., memos, short reports and brief data summaries) and more formal written reports. The evaluation process was observed to be dynamic and ongoing, using evaluation activities and products throughout the evaluation period. Finally, the range and variety of evaluation use recorded during the year's observations suggested a domain of charged use that assumes that use, non-use, and misuse of evaluation processes and products are all realistic forms of evaluation use in real world contexts such as LEA's.

The implications of these findings are that evaluators and decision-makers who wish to use the evaluation process to guide rational change must become increasingly sensitive to the difficulty of conducting and using evaluations in the complex interpersonal and political settings of pressured school district organizations. Skilled evaluators need to learn to recognize when a situation calls for an evaluation "signal" and when instead a more extended evaluation process is in order. Finally, by recognizing the nature of charged use, whether instrumental or persuasive, evaluators should be able to work more effectively with decision-makers to increase use. Considerations for future research conclude the report.

The Process of Evaluation Use in Local Schools

(NIE-G-81-0080)

Abstract

This report presents the results of a yearlong study of the process of evaluation use in a large city school district. Based on extensive naturalistic data-gathering, the research documents how local school administrators use evaluation information generated by the district's research and evaluation unit. The significance of the research is the framework presented for conceptualizing the evaluation use process in an LEA, including the following concepts: three new or revised use factors ("espoused theories" vs. "theories-in-use;" an expanded personal factor; and the clout factor); the types of use (signalling vs. charged use); the evaluation use process; and the domain of charged use. Concluding recommendations are directed toward local school evaluators and users. They emphasize the importance of evaluator sensitivity to organizational complexities within school districts, urging that most time and resources should be spent on evaluations supported by key administrators who have internal clout. Evaluation reports required to "signal" that the ongoing program responsibility is met will be conducted to fulfill external requirements, but these reports may not have the potential to directly effect local program change. Finally, implications for future research, including the call for more naturalistic evaluation use studies, is presented.

Illustrations

	Page
<u>Figures</u>	
1. An Outline of Project Activities	13
2. Types of Information Use: Signalling and Charged Use	37
3. Types of Use: Conceptual Change	46
4. The Process of Evaluation Use	48
5. The Domain of Charged Use: Use (C_1), Non-Use (C_2), and Mis-Use ($\neg C_1$, or $\neg C_2$)	54

Tables

1. The Eight Possible Cases of Charged Evaluation Use	57
---	----

Acknowledgments

A number of people have contributed their time and enthusiasm to this project and deserve more thanks than a brief mention here can provide. These people include the following:

Marv Alkin
Adrienne Bank
Cormell Brooks
Robert Brown
Lois-Ellin Datta
Egon Guba
Frances Harris
Leslie Hendrickson
Freda Holley
Roz Lewy
Yvonna Lincoln
Charles E. Martin
David Ransen
Alan Roecks
Laurie Sabrio
Jo Ann Sanders
Janice Sayas
Nick Smith
Charles Stalford
Bruce Thompson
Suzanne Triplett
Robert Wimpelberg

In addition, we especially appreciate the constant patience and support of Stuart Appelbaum, Cheryl Berry-Gaines, Muriel Rubin, and Fran Smith through many long nights and harried days. Without these encouraging friends and colleagues, this project would never have happened.

Foreword and Rationale

Until recently, the field of educational program evaluation was a growth industry, and school systems and colleges across the country optimistically considered the evaluation process as a vehicle for objective and efficient educational change. By now, however, national economic troubles, coupled with the election of Ronald Reagan as President, have put an end to those early, bright days, and evaluators today, as never before, must demonstrate the value of their wares to users who may be unable to afford such "luxuries" in any case. The irony is that just as evaluation information is more needed to make wisely the difficult decisions of program cutting and retrenchment, local decision-makers' access to such information may increasingly be cut off; evaluation, being a relative new-comer on the practitioners' block, is easily eliminated from already tight budgets at the local level.

Hence, at least part of the rationale for this project is defensive: if we can learn better ways to facilitate evaluation use at the school system level, we may manage to keep alive what we have thus far learned about the process of program evaluation in these settings. To not do this may be to see program evaluation added to education's pile of conceptual has-beens and to see further reductions in R & E staffs nationwide. As evaluators, we feel strongly that this would be both shortsighted and unfortunate.

While we are thus concerned about the potential future of educational evaluation, we are at the same time aware that the strength of the accountability movement nationwide will insure that program evaluations continue in some form. The second part of our rationale therefore confronts use issues

directly. Assuming that program evaluation will in all likelihood exist in local schools, our next purpose is to discover ways to improve this practice, particularly in encouraging local decision-makers to use evaluation information and to use it wisely. That the practice needs improvement is unquestionable; the anecdotal literature of the past ten years teems with negative statements from evaluators and practitioners alike (King, Thompson, & Pechman, 1982).

What can be done to improve use? Studies in recent years have suggested numerous factors affecting use (see, e.g., Alkin, Daillak, & White, 1979, and Alkin, Stecher, & Geiger, 1982), and these certainly deserve study and application in local settings. More importantly in our opinion, however, is the general need for evaluators to wholeheartedly enter their clients' world and, by understanding its reality, to help local practitioners create and use evaluations appropriately.

Consider an analogy. Like the Eloi and the Morlocks in H. G. Wells' Time Machine, evaluators and their clients for the most part live in separate worlds and make only occasional contacts. Though the comparison with Wells' future obviously exaggerates reality, there is enough truth in it to make us as evaluators uncomfortable. Wells, you will remember, presents a distressing forecast in his vision of the Eloi, descendants of the British aristocracy who share a world with the Morlocks, the ever hungry descendants of the working class. In that far distant time, the two communities live harmoniously. The frail Eloi live above the ground and have no unmet needs, either felt or unfelt, while the hirsute Morlocks live below the ground and keep the huge machinery that is their livelihood greased and turning. There is a catch, of course, and a deadly one: at night the Morlocks climb from their subterranean lairs to feed on the childlike and defenseless Eloi. What at first appears to the Time Traveller a paradise reveals itself finally to be a continual nightmare for its inhabitants.

To administrators and other practitioners, evaluators may seem like the Eloi--naive, spoiled, out of shape, and used to getting their way--while they, like the Morlocks, toil in the stifling reality and overpowering constraints of their school settings. If it is remembered that the Morlocks devour the Eloi one at a time, the image of the hapless evaluator in the throes of a controversial study can serve as a reminder that the job of evaluator is not without its dangers.

To evaluators, on the other hand, decision-makers and other clients may seem a bit like the Eloi. Unappreciative and even frightened of the impressive social science machinery the evaluator controls, they live in a world apart from the rigors of multiple regression and factor analysis. Unless dragged, they rarely enter the evaluator's world, preferring the bliss of ignorance to so-called "objective" realities, and yet not hesitating to use the empiricist Morlock products when these suit their needs.

We have purposely exaggerated the separation between evaluators and users because--like it or not--we must acknowledge the fact that the only place our work can succeed is in the local users' world. If our results are not used there, they are not used at all; and if they are not used at all, someone--probably us--has wasted an enormous amount of energy in an era of severe energy shortages. Our message reduces at last to this thought: the world of the evaluator must not merely become like the users' world, but it must be the users' world. We must not only enter the world of our local clients; we need to hang out a shingle and share the office space.

Such reasoning, then, provides the further rationale for the study to be described in this report. Not only should the current study help us demonstrate the potential merit of program evaluations for local users; it should ultimately show us practical ways to make such evaluations worth their cost to LEA's.

This report is the product of an NIE small grant awarded to the Orleans Parish School Board (NIE-G-81-0090), the second such grant awarded. The first (NIE-G-80-0082), in 1980-1981, generated five products: an extensive bibliography of the evaluation use literature (King, Thompson, & Pechman, 1981a); a shorter, annotated bibliography (King, Thompson, & Pechman, 1981b); a review of the literature (King & Thompson, in press b; King, Thompson & Pechman, 1982); a nationwide survey (King & Thompson, in press a); and an empirical simulation study (Thompson & King, 1981).

Based on this initial work, a second grant was written and funded for 1981-1982 to study in depth, using largely qualitative methods, the process of evaluation use in a large city local educational agency (LEA). The methodological basis of the study comes from the notion of grounded theory (Glaser & Strauss, 1967), i.e., the use concepts we sought originated directly in the field data we collected. For this reason and because we were funded for just one year--a short period for any naturalistic work--we opted continually during the year to collect as much information as possible and to spend relatively less time reporting extensive process analyses of the results.

This year-end report of the NIE-funded study, then, represents a first effort to integrate all of the data collected. It is not so much a "final" statement as an initial exploration of ideas and concepts stimulated by the year of research. The report will be subject to revisions over the course of the coming year. During that time, we will also integrate the existing literature on the topic; here we cite only obvious pieces and omit many that deserve mention. Readers of this document are therefore encouraged to check with us for additional papers and analyses based on these results (c/o Dr. E. M. Pechman, Director of Testing and Evaluation, Orleans Parish School Board, 4100 Touro Street, New Orleans, Louisiana 70122). Products will

ultimately include a set of thick descriptions of the ten or so cited case studies, a set of workshop materials for training administrators about use, and an extensive report of the evaluation use literature in light of these findings.

Section 1: Introduction

In the boom days of the '60s and '70s, evaluators could afford to be an optimistic lot. The impetus provided by linking federal monies to program evaluations, together with a growing sense in the field of just how such evaluations could aid decision-makers in varied settings, led to the hope that a rational and well-tempered approach to education could succeed. The image of the overworked decision-maker calmly sitting to consider evaluation information before taking appropriate action could bring a smile to the evaluation community's face.

But belief in this image faded quickly as evaluators who had done everything by the book saw their efforts disregarded and their well-done reports, time and time again, placed in obscure filing cabinets, never again to see the light of day, or ignominiously consigned to gather dust among rarely used books. The initial literature on evaluation use abounds with gloomy perceptions of the "failure of educational evaluation" (King, Thompson, & Pechman, 1982). At one point, Robert Stake went so far as to wonder whether "evaluation is going to contribute more to the problems of education or more to the solutions" (1976, p. 1). This perspective is what Alkin, Daillak, and White call the "mainstream" viewpoint of evaluation use, i.e., the viewpoint that "contends that evaluations seldom influence program decision-makers and holds out little hope that evaluation will ever break through the barriers to real impact on programs" (1979, p. 17).

The results of empirical studies of use, however, have challenged the worst-case perceptions of this mainstream perspective, and an "alternative" viewpoint has recently emerged (Alkin, Daillak, & White, 1979). Studies by Patton et al. (1977), Alkin, Daillak, and White (1979), Dickey (1980), and

Kennedy, Apling, and Neumann (1980) suggest that evaluation results are used by decision-makers, "but not in the clear-cut and organization-shaking ways that social scientists sometimes believe research should be used" (Patton et al., 1977, p. 144). The current view of evaluation use, as summarized by Brown and Braskamp, is that

. . . effective utilization does not necessarily mean that any of the recommendations are implemented or that there are any immediately apparent decisions based on the information (1980, p. 91).

Rather, results frequently influence users in indirect or gradual ways, suggesting that the discouraging mainstream picture of evaluation use may have stemmed more from a mistaken expectation regarding the nature of use than from the actual level of use in the real world. Wise (1978) has written that

If there is an evaluation utilization problem, it is not that decision-makers do not use the information they receive; it is that evaluators cannot easily see their information being used in the incrementalism of real-world decision-making (p. 24).

Or, as Pogo so eloquently once said, "We have met the enemy, and they are us." No longer, then, is it sufficient to look for direct and readily apparent instances of use. Supported by empirical data and common sense, the alternative viewpoint has challenged the traditional; and we can now assume that use occurs at the local level, although it is largely undocumented and little understood.

An important and parallel development in the study of evaluation use has accompanied this conceptual shift. Many studies to date (e.g., Patton et al., 1977; Dickey, 1980) focus on the use of evaluation products, typically the final reports of federal projects. This approach retrospectively traces the ways in which decision-makers have considered and applied the results of evaluation reports once received. To examine the process of use from this perspective, then, is to look at how people have used, or say

they have used, existing reports. The appropriateness of such an approach for the initial studies of use is unquestionable. The mandated evaluation products associated with federal projects insured researchers that practitioners did, in fact, have something to use.

However, an important limitation of this research must also be noted. As other studies have shown (e.g., Daillak, 1980), to focus solely on the use of the end-products of evaluations--and especially on externally mandated end-products--is to ignore the complex process involved in the creation and ongoing use of evaluation information in a given context, the very process suggested by the alternative notion of use. The dynamics of evaluations are rarely simple, clearly changing over the course of the evaluation as events occurring early in the process influence later actions. The ultimate and remembered use of a final report is just one piece of an intricate puzzle.

The alternative approach to studying evaluation use is therefore neither product-centered nor retrospective. Because use can occur in many ways throughout an evaluation, this method looks not only at products, but at as many of the events and interactions leading to them as possible. Whereas the first approach examines the final use of evaluation products, the other looks instead at the evaluation behavior of both evaluators and decision-makers, not excluding their use of products, but including much more of the informal, persuasive, and political interactions involved. To capture the ongoing, context-centered nature of evaluation use, this approach necessitates a prospective orientation quite different from the first.

The study described in this report uses this second approach to address two questions: (1) In what ways do local decision-makers use evaluation information generated by the research and evaluation unit of a large public

school system? and (2) What concepts can describe the use of such information within a local educational agency (LEA)? Following a description of the research setting and methodology, the report is divided into three sections: first, a discussion of the assumptions about local evaluation use that are questioned by this study and a proposal for rethinking evaluation use in local settings; second, a presentation of the concepts generated by our data; and third, a summary of the implications this information has both for the conduct of program evaluations in school settings and for the conduct of related research.

Section 2: The Research Setting and Methodology

Because of our earlier review of the evaluation use literature (King, Thompson, & Pechman, 1982), the choice of methodology was, for us, an obvious one. As Alkin has consistently argued:

The forces which lead to utilization are indeed complex. This complexity in combination with our current inadequate understanding of evaluation and utilization requires a methodological procedure sufficiently sensitive to capture the nuances involved--naturalistic research is currently a most appropriate tool for a study of evaluation utilization (1979, p. 13).

The difficulty of generalizing from a single case study notwithstanding, only naturalistic methods can trace, record, and add into theory the many critical variables simultaneously present in a complex evaluation setting (Glaser & Strauss, 1967). This point is well documented by the earlier case studies of Alkin, Daillak, and White (1979) and Daillak (1980); and we sought to extend the use literature through naturalistic study of the evaluation use process occurring in a large city school system, starting with the

activities of a research and evaluation unit* and then following examples of use as they developed.

The selection of this site was highly appropriate for three reasons. First, the system faces all the problems characteristic of modern urban education. It is a large city public school system of over 80,000, with a heavily minority student population. It is run by an elected board, an appointed superintendent, and a highly centralized bureaucracy. Like all urban districts, it has received substantial federal funding through ESEA Title I and IV-C (14% of the budget for FY 1981), but for a variety of reasons there is little money available in the system for even basic supplies and capital maintenance. Despite the hard work of many people and their numerous successes, the system's reputation in the community is not encouraging, a fact that infuriates some employees and constituents and discourages others.

A second reason the system was appropriate for our study is that since 1975 it has had a separate R & E unit charged with maintaining the student data base, completing essential state enrollment reports, running the system-wide testing program, and conducting required educational program evaluations. During the year of the study, the department had a staff of 18, including two full-time evaluators, two part-time research analysts, and a director with a commitment to making program evaluation responsive to the system's program needs. In such a setting, one would expect either to see the use process in action or to discover the factors inhibiting local use.

The third reason for the selection of our study's site was pragmatic: we were fortunate to have both access to the R & E unit in question and

*Hereafter referred to as the "R & E unit."

their support for our project. With interest and some skepticism, the department members were receptive to participation in this study, and we had open access both to the staff and its processes. Also, a letter of support from the superintendent enabled us to observe and interview throughout the system for one complete school year.

This access and support did not come without certain costs. Rather than conducting a relatively pure case study as we had originally intended, we found ourselves over the course of the year serving as participant-observers in a departmental self-study. The clearest problem in assuming such roles was that of potential contamination and co-optation, i.e., of losing the "objectivity" deemed essential by the scientific method. Of this we are, no doubt, at least partly guilty.

In retrospect, however, the benefits to the study in our opinion clearly outweigh any costs, because the distinct roles assumed by the four project staff members afforded us many dynamic and varied data-collection opportunities. These roles were as follows:

1. The project director of the grant was the department's director who chose to open herself and her department's activities to analysis in the belief that the local school evaluation process is highly complex and little understood. She succeeded in persuading her staff of the direct and indirect value of a self-study analysis.
2. The first author of this report was an outside participant-observer, responsible for attending important meetings, conducting major interviews, and balancing the director's "insider" viewpoint with that of an "outsider's" objectivity.
3. One person began as a research assistant, tracking instances of use. In the middle of the project year, she was hired part-time in the R & E unit's testing office and lived the life of an R & E staff person for several months.
4. One person began and remained an outsider to the department, responsible for conducting numerous personal telephone interviews with key individuals around the system.

Frequent meetings of project staff allowed us to compare notes and share information from our differing sources and perspectives, leading us ultimately to the ideas presented in this report.

Participating in the R & E unit's activities while observing allowed us to experience evaluation use firsthand; indeed, we were pro-actively part of the ongoing use process, witness to its occasional triumphs and to its more frequent frustrations. Our year-long working relationship with R & E unit members and other system employees also helped to develop mutual trust and understanding and to minimize problems that might have otherwise existed. If we are therefore suspected of somewhat less than ideal objectivity in our data collection, we would defend ourselves on two grounds: first, that the data we have is far richer and more accurate than we might otherwise have obtained; and second, that the concept-generating goal of this study makes this less a problem than it might otherwise be. Additional research will be needed to validate the ideas presented here, our purpose being to generate concepts, not to test them.

Our project consisted of two phases (See Figure 1). Applying Guba's (1978) wave conceptualization of inquiry, the first phase was an intensive self-study of the research and evaluation unit of a large city school district. To use Guba's terminology, this phase moved between discovery and verification modes and consisted of two "waves." The purpose of the first wave was to create a portrait of the workings of the research and evaluation unit as seen by its staff, i.e., by those directly involved in data collection and potential use. We began by conducting extensive interviews with all 18 R & E staff members (evaluators, clerks, administrative assistants, etc.), with external evaluators who contracted with the department, and with the key school system staff who interacted regularly with evaluation person-

Figure 1. An Outline of Project Activities

Phase 1: Intensive Self-Study of an R & E Unit

"Wave"	Purpose	Activities
1.	To create a self-portrait of the R & E unit.	Interviewed R & E staff, external evaluators, and key system staff who interact with unit; observed in offices; examined documents.
2.	To create a conceptual framework for the process of evaluation use in the system as a whole.	Interviewed top management; tracked R & E's major evaluation projects; "chased" ten use instances; interviewed (by telephone) 20 key users; continued observing and reading documents.

Phase 2: Revision of the Conceptual Framework

Activities

Sent draft of conceptual framework to R & E directors and evaluation use researchers.

Revised conceptual framework based on their reactions and comments.

nel. The outside participant-observers also began to observe day-to-day life in the office--attending meetings, conducting on-site visits, completing data forms, interacting with the system's administrative structure, and so forth. In addition, we examined available documents and reports so that the dynamic structure of the evaluation context began to emerge. The write-up of the description of the unit's functioning and the related member check concluded the first wave of the study's initial phase (See Pechman & King, 1982).

The second wave moved from the internal perceptions of the R & E unit into the context in which evaluation use occurs. Following our initial observations and during most of the year, we worked simultaneously on five interrelated activities:

1. We conducted personal interviews with top management decision-makers (i.e., the superintendent and various deputy, assistant, and district superintendents) to determine their views about the nature of an ideal R & E unit and their perceptions of the reason for any disparity between the department's achievements and their ideal. We finally discussed their part, if any, in specific on-going evaluation projects.
2. In order to gain a sense of the process involved in generating and using evaluation information, we closely followed the progress of two of the R & E unit's major projects, one the mandated evaluation of a federally funded program and the other the superintendent-requested evaluation of a costly staff development and curriculum improvement program. For the first program, extensive interviews were conducted with the federal project staff, with an interested school board member, and with the R & E staff (the director, the evaluator currently assigned to the program, and the

former project evaluator). In the second program, our role became that of observer-participant, attending meetings, informally talking to some participants, formally interviewing others, and even strategizing with the R & E staff regarding approaches to the program's evaluation.

3. In the first wave, we had identified ten cases where evaluation use seemed to (or was reported to) be occurring. We called this data gathering process "instance chasing" because, for these cases, we closely examined each major aspect of the evaluation process: we read available, relevant written materials; we reviewed evaluation activities; and we interviewed and re-interviewed the primary and secondary users, the assigned evaluator, and others who were or should have been involved in the evaluation. These completed instance studies, representing among others a migrant program, a program for promoting the use of educational television, a program for highly-motivated bidialectical black children, and a study of first grade achievement in high need, reduced sized classes, helped provide extensive--some might say confusing--data about the process of local evaluation use.
4. The fourth activity was to conduct 20-30 minute telephone interviews with key users who had not been interviewed in person, but who either were nominated by an interviewee as a good source of information or who, for whatever reason, were referred to repeatedly in our discussions with systemwide personnel. This group included about 20 project coordinators, central office staff, and principals who were asked questions created from categories derived from our initial interviews.

5. In addition, project staff continued to serve as both participants and observers of R & E activities and to collect and read relevant materials. It was at this point that one research assistant became a part-time R & E staff member.

The completion of these activities concluded our data collection, and the first phase of the project was then completed with an analysis of available data. These data included the following: notes from interviews with roughly 80 people, eleven of whom had been formally interviewed twice; notes from meetings attended and from other observations and discussions in the R & E office; evaluation memos and reports from seven federal programs, four locally sponsored evaluations, and four other R & E projects; past years' office files on R & E's two current major evaluations; notes from our staff meetings during the year; and the several papers we had written in conjunction with the project. In addition, some of the data collected in the system during the year were used in other evaluation-related studies (Alkin, Stecher, & Geiger, 1982; and Rubin-Frankel, 1982).

Recognizing the limits of studying a single case, the second phase of the project integrated the evaluation use concepts developed from the data with the perceptions of evaluation researchers and practitioners. A summary of the study and the derived conceptual framework was sent to a sample of large city R & E directors, evaluation theorists, and other evaluation practitioners along with a reaction questionnaire. The responses were analyzed and the study's concepts then revised as necessary, thereby completing the theory generation.

A follow-up activity planned for spring, 1983, will be the administration of a systemwide questionnaire to instructional supervisors and coordinators, local principals, and a random sample of 500 teachers in an attempt to begin validating the conceptual framework developed. At that time, we may also conduct a nationwide survey for the same purpose.

Section 3: Rethinking Local Evaluation Use

Questioning Earlier Assumptions About Use

Recent literature has labelled two earlier assumptions about evaluation use as myths. The first is the "big bang" myth, which maintained that the results of evaluations would create instrumental use in its classical sense, i.e., immediate and observable effects in local settings. As noted in the introduction, such use is not that common, and the corollary "mainstream" viewpoint that evaluations are rarely used has also been replaced by the alternative viewpoint that decision-makers do use results, although in more subtle ways appropriate to the organizations in which they live and work (Alkin, Daillak, & White, 1979; Kennedy, 1982). As our data repeatedly show, we are more likely to find "small whimpers" of use in local settings than big bangs.

A second such assumption now viewed as myth is that of the "rational decision-maker," i.e., of the existence of a solitary, classically rational decision-maker or, more generally, of the possibility of such rational decision-making in organizations. In its fullest version, the myth goes like this:

Harried decision-makers, overworked but rational at heart, are presented evaluative information that is relevant to decisions they are about to make. Sitting calmly at their desks, they consider the data, weigh their options, then make the "correct" decision, i.e., the choice supported by the data. The evaluator, beaming happily behind the scenes, takes pride in nurturing such rational action (King, Thompson, & Pechman, 1982, p. 4).

Other literature has discussed the reasons for the unlikelihood of such a scene's occurrence (e.g., Weiss, 1979; Cronbach et al., 1980; King, Thompson & Pechman, 1982), and the myth is now seen as having at least two related and false assumptions: 1) the possibility of a single and correct "rationality" for an organization as a whole; and 2) the idea of a single

decision-maker making decisions apart from the political context in which he finds himself. Instead, rationality must be recognized as a relative notion--the "right" of each individual--although to do this greatly complicates the search for evaluation use. Nevertheless, the data collected this year clearly support this notion; multiple perspectives and rationalities are a way of life in any LEA bureaucracy.

If, then, the debunking of the big bang and the rational decision-maker assumptions are supported by our data, what about other commonly accepted assumptions about the use of evaluation information? Last year's review of the literature pointed to two additional assumptions about how to encourage evaluation use: improving the quality or catchiness of evaluation reports; and ongoing collaboration with potential users throughout the evaluation process (King, Thompson & Pechman, 1982). This year's data speak in part to the first assumption and more thoroughly to the second, suggesting that these solutions may be partial at best and, like the earlier myths, may appeal more to our sense of the way things should be than to an awareness of how things are.

Improving the quality of evaluation reports. Evaluation reports have received bad press in recent years. Denny (1980), for example, notes that "Most of these reports have one omnipresent quality: they are dull, dull, dull" (p. 4). Datta (1979) reports that "Although titled 'evaluations,' more accurately . . . [the reports she examined] might be presented as descriptive statistical accounts of some aspects of educational programs" (p. 17). And Cronbach et al. (1980) write that because the "evaluator's final, formal report is essentially an archival document, not a live communication" (p. 185), the likelihood of its having impact will in most cases be small.

Although any recommendation for improved reporting is at best only a partial solution to the problem of evaluation use, it is important to recognize the potential danger of ignoring such reports, especially since they can be important sources of information for busy decision-makers in local settings (King & Thompson, in press a). Evaluation writers have therefore suggested two ways to improve reporting. The first includes ways to improve written reports, for example, by making the evaluation report a "well told story," including details of the evaluator's personality, e.g., his competence, style, beliefs (Denny, 1980, pp. 5-6); or by using executive summaries on colored paper to report major findings, and appendices to present more technical information (Alkin, 1975, pp. 208-209). Datta (1979) also recommends the use of executive summaries, i.e., that evaluators "address the few important questions, present the evidence, and state the action implications in one page or less" (p. 23). The second approach to improved reporting, advocated vigorously by Freda Holley (see, e.g., Lee & Holley, 1978), employs varied and more effective media, increased informal contacts, and organized press releases to get evaluation information across to possible users.

In our observations we witnessed no examples of this second approach to evaluation reporting. We did, however, examine cases where the first approach was followed, but to no avail. In one large federal project evaluation, for example, the evaluators succeeded in developing the mandated state report into several potentially useful formats: a lengthy statistical compilation for the State Department of Education; a more detailed technical volume for the staff; and individual community summaries for each school. The various reports effectively communicated evaluation results in a specialized manner for each audience, and they were both formally and informally presented. But fine reports in no way were able to overcome the basic

constraints of the highly political evaluation context. Faced with negative gain scores, the federal project director ultimately challenged the data base, and the evaluation information was never disseminated or used.

A second example in which a good report went unused occurred when a superintendent headed a committee studying the question of whether or not the system should invest in extensive, new, computer-based innovations. Interviewed a short time after the carefully written results and analysis of a survey were hand-delivered to him, he could think of no instance where the R & E unit had provided him with useful information. Although he and his committee reviewed the report's contents, a newly appointed superintendent simultaneously took charge of the planning committee, and the survey results were simply disregarded.

If, then, on the one hand, good written reports did not necessarily lead to use, on the other, poor reports did not inevitably result in non-use. In one instance an evaluator, recognizing that a user wanted only the simplest frequency counts to make an argument to support her program, gave up on trying to encourage her to use more fully analyzed data that would have given better information to act on. In another case, a simplistic and distorted ranking of schools' state test results was used to begin the year's curriculum planning, in spite of the availability of other comprehensive data that presented a more informative but complex picture of schools' needs. The latter report failed to point directly to one or two needy schools and was rejected as too complicated and open to various interpretations.

What can finally be said about the effects of improving the quality of evaluation reports? Our sense is that there is not necessarily a direct relationship between improved written reports and improved levels of use. If a user finds written information of value to his particular situation, he is

likely to use it, regardless of its quality; if, for whatever reason, it does not seem of immediate value, he will not use it. Of course, evaluators should still strive to write the best reports possible. That, after all, is part of our job. But the intrinsic power of any well written report, in and of itself, must be recognized as trivial in comparison with the additional, powerful forces working in an evaluation context. Other literature suggests that the same is true for the methodological quality of the evaluation itself (King, Thompson & Pechman, 1982).

Ongoing collaboration with potential users. King, Thompson & Pechman (1982) provide numerous citations suggesting the importance of informal contacts and ongoing collaboration between evaluators and decision-makers for increasing the use of results. The assumption seems appropriate; to the extent that an evaluator focuses the evaluation on information of interest and value to the decision-maker, the results should be usable and, presumably, used. As was the case with writing good reports, to not do this is, in some sense, to abrogate responsibility for conducting quality evaluations. However, our data again force us to question this assumption as they point to cases where sincere attempts at collaboration did not lead to improved use.

The case of one federal project is a good example of this. The evaluators set out to conduct the best evaluation they knew how, one that would go far beyond the state requirements to be both methodologically sound and of interest and value to the project staff. Repeated memos for over a year document the willingness of the R & E staff to meet with project people to insure that the process was meeting the program's needs.

The initial problems with the collaboration stemmed from an effort to obtain accurate enrollment information about the program. Merely by attempting to get basic data about each school participating in the project,

the evaluators inadvertently stepped on several toes: project staff resented a perceived intrusion into what they considered their domain; school personnel resented the extra work and suspected the real reasons for the inquiry; and Computer Services resented having to reprogram their basic data files. It is unlikely that these problems could have been avoided because the project was so politically sensitive that the nature of the situation made outsiders to the project suspect. Once on guard, the project staff effectively thwarted efforts at joint work. The data eventually gathered for program planning was never even distributed beyond top project staff.

A second example of attempted but ineffective collaboration occurred when a new superintendent sought an evaluation of an existing staff development and curriculum improvement program. She asked for the answer to a question she regarded as simple--Was the program effective?--and became frustrated and irritated when the R & E staff involved her in more explicitly defining the evaluation questions, i.e., in deciding what type of information she would consider as evidence of the program's success. She viewed R & E's repeated contacts as evidence of uncertainty on their part; as evaluation professionals, she expected them to make all decisions related to the evaluation and not to "bother" her until the final go-no go data were in hand.

At another level, program coordinators also balked at becoming involved in the evaluation. Meeting after meeting was held to define evaluation questions of use to them, questions that would accurately reflect the good they felt the program had done and help them keep their jobs. Perhaps sensing the hopelessness of their position, they turned the meetings into gripe sessions and were only able to make decisions when the R & E staff gave them limited options to be rank ordered or voted on. The controversial

history of the program coupled with its likely fate made collaboration all but impossible.

In a third example, the collaboration between evaluators and users--the heads of several minor federal curriculum projects--worked well but to no avail. In these cases, the information collected simultaneously met state requirements and documented the projects' successes in establishing strong curricular innovations on shoe-string budgets. The results of these effective collaborations were, however, of little value because outside of the project staff no one cared enough to act on the evaluators' recommendations to extend the programs beyond the federal funding period.

These examples point to the practical difficulties of evaluation collaboration in an LEA. In the first case, the political factors surrounding the project prevented trust or collaboration from developing. In the second, the potential users considered collaboration unnecessary and an indication of poor management, while in the third, the collaboration succeeded, but the hoped-for outcomes still did not ensue. Although we can also point to collaborations that did lead to instances of use, attempting cooperative action in and of itself may not provide a sufficient condition for evaluation use.¹

The Futility of Finger Pointing

The conventional expectations about program evaluation discussed previously smacked highly of common sense. As the field developed in the 60's, it somehow made sense to expect that decision-makers would act "rationally"

¹ Nor, technically, is it necessary. An interested user can certainly "charge" evaluation information and create use on his own, with or without any evaluator assistance (see Section 4).

in their day-to-day activities and take advantage of the information evaluators created for them. Almost twenty years of experience, however, have demonstrated the naivete of this expectation. Similarly, it made sense to expect that evaluation results, communicated in the most effective ways possible, should improve use, and that evaluators who collaborate with the ultimate users of their information will see their processes and products applied. While our data indicated some instances where this was the case and the expectations were happily met, examples given previously also point to the breakdown of these assumptions.

The question then arises as to why these assumptions about improving use do not always appear to hold true. One answer to this question has been to lay the blame for such failures on a specific group of individuals, either users or evaluators. Just as pin-the-tail-on-the-donkey has entertained two or three generations of birthday party goers, pin-the-blame-on-someone has been a popular pastime for evaluation theorists in recent years. At first the blame was pinned squarely on local users. Who were they, after all, to ignore the implications and dictates of finely crafted evaluation reports? The evaluator's job was to prepare these reports; the user's job was to make use of the small pieces of truth objectively doled out. If such information went unused, it was clearly their fault. The relation of this position to the big bang and rational decision-maker myths is obvious.

As comforting as such an attitude was, however, some writers turned to Caplan's two community theory (Caplan, 1979) noting that the "scientific" approach of many evaluators did not mesh with the more subjective reality of local users. These writers (e.g., Wise, 1978; Cronbach et al., 1980) for-

gave the user and instead blamed non-use on the evaluator, giving emphasis to the role of evaluator as "educator." If users ignored evaluation results, it was because the evaluator had not communicated with them effectively or, in the evaluation process, had failed to deal with the complexities of the evaluation context. "Mea culpa," shouted many in the evaluation community, and the search for new ways to generate use was on.

While there is no doubt that guilt and finger pointing can be therapeutic, the data collected this year suggest that the problem is not simply one of placing blame or responsibility. The examples discussed in the previous section on collaboration demonstrate that in some situations evaluators are doomed to fail, regardless of their methodological skill and interpersonal finesse. In highly politicized settings, merely collecting the most basic data, to say nothing of disseminating the results, may literally be impossible; as we have previously demonstrated, a few individual actors can insure use or close off its possibility. The entire notion of the evaluability assessment appropriately allows for this (Rutman, 1980), but such assessment may not be a realistic alternative in the product-oriented school bureaucratic environment.

But just as non-use is not always the evaluator's fault, our examples provide evidence that neither is it necessarily the fault of the user. The evaluations of one well-received music project, for example, repeatedly gave the concept and the staff high marks. When its funding ended, however, the system allowed the program to die, despite the supportive lobbying efforts of both the parents and teachers in the school whose students had been involved. Given the constraints of budget cut-backs, the money to continue the program was simply not available, and to assign blame for not using positive evaluation results misses the point. A second case is that of a federal project head whose evaluation urged additional clerical support. In

no way could she apply the results and get a typist because the distribution of clerical staff was beyond her jurisdiction. For whatever reasons, her superior "used" the evaluation report only as a basis for a lengthy letter of reprimand; it did not lead him to provide the needed help. But is it appropriate to blame him for not "using" the results completely or "correctly"? Individuals in LEA settings act not only as individuals, but as institutional representatives. The essays on school organizations in Bank and Williams (1981) provide ample support for the institutional dynamics that are part and parcel of the evaluation process. Users would be naive to ignore such pressures and do only what they personally choose to do.

The point is this: While in some cases the reason for non-use can be attributed to failures on the part of a specific individual, whether user or evaluator, in many others it cannot be. As King and Thompson (in press b) and Smith (1982) note, the relation between evaluator-user and use parallels that of teacher-student and learning. The responsibility must, in the final analysis, be shared. The question then becomes one not of assigning blame, but of finding ways to encourage a sense of shared responsibility. Achieving this sense of involvement is not just the evaluator's problem; it is also the user's because the nature of evaluation requires the user to participate actively in the process. Sometimes this means deciding not to conduct an evaluation (or to conduct a limited data collection effort only to complete the requisite forms) due to the obvious constraints of the evaluation setting.²

² While this discussion clearly ignores the issue of how to effectively involve multiple decision-makers and audiences, that topic is beyond the scope of the current report.

Additional Factors Affecting Use

Previous studies have compiled an extensive list of the factors affecting evaluation use (see Alkin, Daillak, & White, 1979; Leviton & Hughes, 1979; King, Thompson, & Pechman, 1982; Alkin, Stecher, & Geiger, 1982). To date our data suggest three additions or revisions to this list: 1) users' "espoused theories" vs. their "theories-in-use"; 2) the personal factor reconsidered; and 3) the clout factor. Two of these factors concern the characteristics of the local user and the third the nature of the local evaluation context.

"Espoused theories" vs. "theories-in-use." Just as the evaluation literature's assumptions for improving use need examination in light of case study evidence, so, also, do the evaluation assumptions of local users. In our interviews and discussions with users this year, three ideas repeatedly surfaced to suggest that "espoused theories"--what people say they believe--may differ dramatically from their "theories-in-use"--what their actions imply they believe (Argyris & Schon, 1974; see also Kennedy's concept of "working knowledge," 1982). Whether this was the "good subject" effect, i.e., people telling us what they thought we wanted to hear, or some deeper belief in "rationality" is a point in need of further study. These espoused theories were the following:

1. "I believe in evaluation." Without exception, each person we talked with made this point strongly, although what exactly they meant by "evaluation" was often unclear. For some it was filling out the forms related to accountability and compliance. For others, and particularly for one federal project head, it was a visible threat to keep program people "on their toes." Many saw evaluation information as documentation--of program implementation; of program success (the "good news"); or of program failure

(the ammunition for forcing change). Still others thought of evaluation as the computer print-outs of achievement test results to quarterly summaries of basic program data such as staff size, number of children served, and documentations of meetings held. Outside of the evaluators, very few saw the purpose of evaluation as the judgment of "merit and worth" (Guba & Lincoln, 1981).

If this was the espoused theory, the theory-in-use was frequently one that made effective evaluation all but impossible. One superintendent, for example, complained that the R & E unit was too expensive, pointing back to an earlier time when one woman was able to "do the work of the entire unit" and, in his opinion, to do it better. Other project heads said they believed in evaluation, but either worked against the evaluators during data collection or took an inactive role that in the end helped to undermine the process.

2. "Evaluation is helpful to programs." While the rational decision-maker myth has been questioned by those responsible for conducting evaluations, a version of it lives on in the comments of local users. Repeatedly we were told that the results of the evaluations we examined would be of value in improving the programs, even when it was clear that the very process was creating tensions difficult to resolve and that people were unnecessarily complicating or thwarting the process for personal reasons, indicating a different theory in use.

Two examples suggest this espoused theory. The first is a curriculum improvement program that was eliminated on the basis of interim evaluation data. During seven months of meetings, a frequent comment made by administrators and evaluators was that the evaluation would point out the strengths of the program so that these could be rebuilt into something better. This became something of a chorus. Meanwhile, the superintendent in

charge disregarded what these strengths might be, and the teacher-coordinators were so worried about losing their jobs that many felt unable to do their work for much of the year. The difficulty of conducting an evaluation involving these users led the evaluator, new to the system, to comment that, "When you're outside this building, you assume that decisions are made with some sort of rationality. Then when you get inside, you realize it's not true." While in one sense the evaluation was ultimately "helpful"--it eliminated a costly, poorly coordinated program--in another it was highly dysfunctional: people were unable to function effectively during the last year of the program and, in the end, the coordinators did lose their non-classroom jobs.

The second example is the federal project that the R & E unit attempted to support by providing an excellent, audience-focused evaluation. Again, everyone involved wrote memos stating that the evaluation would be of use to the program, but in the end a series of unsubstantiated accusations questioning the basic evaluation data led to the complete breakdown of the collaboration. The time spent by both program people and R & E staff arguing about the evaluation procedures indicates that the evaluation was not in any way helpful to the project, and instead caused considerable unresolved tensions that continue to this day.

3. "It's not hard to do an evaluation; you just get out there and do it." This statement--what we call the Rumpelstiltskin effect--was made by virtually every top administrator we talked with in the system and by several project directors and teachers. These users were distressed by the R & E unit's presumed inability to "get on with their job," by their repeated harping on the difficulties of collecting useful data, and by their continuous concern with being adequately funded and staffed for what they needed to do. The administrators expected evaluators to weave straw into

gold and to work data-based miracles, all with no special support, either financial or nominal.

One superintendent explained how he would go about conducting an evaluation: "Their (R & E's) hands aren't tied...Evaluators can communicate with program people. You just tell them, 'I need your help,' and they will respond." The assumption is that obtaining "objective" information to accurately reflect educational program status and effects is easy. This, of course, ignores the entire political context of the school bureaucracy as well as the reactions of individuals to the evaluation process. So, while the theory-in-use helps thwart effective evaluation by ignoring the problem, the espoused theory makes users incredulous at what they see as the continued failure of the R & E unit to do what the system requires them to do.

The personal factor reconsidered. Patton et al. (1977, p. 64) define the personal factor as the presence of an identifiable individual or group of people who personally care about the evaluation and the information it generates. The data collected this year point to characteristics that describe the type of person likely to care about the use of such information. It seems that self-confidence (both in a personal and programmatic sense) and interest in and openness to evaluation information separate likely users of evaluation information from non-users. This suggests an important and, in many cases, uncontrollable factor affecting the use of evaluation results because openness to evaluation may, in some instances, be politically foolhearty and nonproductive to the achievement of a decision-maker's longer term goals.

The easiest way to see this effect is to compare instances of use with those of non-use. In those cases where information was used, the project directors or administrators typically believed in their programs, i.e., knew that what they were doing was important for students, but more importantly, they felt that the improvement resulting from an objective evaluation proce-

duration would increase the chances of the program's survival in the system. Their investment was not in direct personal power or control, but rather in doing a job well. At the same time, the evaluation process functioned only where the political context allowed the program director to open the program to review. In one instance, a lower level coordinator of an innovative early reading program very much wanted a candid and useful evaluation. However, one of her superiors didn't place much value in the program and this reality reduced the evaluation process to a routine and simplistic affair. The disinterested superior controlled access to the program, leaving its evaluation-oriented coordinator with no other choice than to limit the evaluator's involvement in data collection.

Thus, the evaluation process succeeded when key staff understood how evaluation helps programs, when they were open to information from any source, and when the situational politics were favorable. Evaluation proponents were not frightened by bad news, were looking for successful innovations and new ideas, and were in a position to create conditions that enabled the evaluation process to occur.

We initially thought that the self-confidence aspect of the personal factor might have something to do with program advocacy, i.e., that people who truly believed in a program would be sincerely seeking ways to insure its success and that evaluation use would be one part of doing this. One federal project, however, demonstrated clearly that an administrator could believe in a project, but, in pursuit of longer term goals, work against an evaluation that might have suggested improvements. On the other hand, another administrator, who did not strongly advocate the program being evaluated (although recognizing its potential strengths), was open to any information suggesting good ideas that could be worked into existing or new programs. Program advocacy, then, seems to be neither necessary nor sufficient for the personal factor to emerge. Instead, it is a personality type

in combination with a favorable political context and climate, a self-confident and open person able to take the good with the bad who seems to be able to use evaluation to the best advantage.

The clout factor. The third factor we would add to the list of those affecting evaluation use is a new one. The clout factor refers to support for the evaluation process by critical members of the top management hierarchy. This support can take two forms: funding for the actual cost of evaluations, either on a study-by-study basis or for the R & E unit as a whole; and, equally important, direct and visible administrative involvement. For example, evaluation generally succeeded when it was supported by written memos from key administrators encouraging or mandating others' participation. Decision leaders' attendance at significant meetings and their willingness to speak favorably about the importance of the evaluation process at regular staff meetings and in their ongoing interaction with program staff also increased the likelihood the evaluation process would succeed. Our data suggest that the political reality of the LEA context requires the explicit and continual action-based support of powerful administrators; without it, evaluators may struggle in vain to convince people first to cooperate in the evaluation and, second, to use its results.

The point is not, however, as simple as "If you have clout, your evaluation results will be used, and if you don't, they won't." Because clout is applied in a complex and often highly political setting, its effect will vary. The discussion of the personal factor suggests that some people will both seek and use evaluation information because it is personally important to them to do so, regardless of what their superiors urge them to do. A second case where local clout may have little effect is that of institutionalized evaluations, for example, those tied to federal funds. Knowing that funding requirements mandate that evaluation reports be returned, even

if no one "uses" them locally (see Section 4 for a discussion of this point), people will participate in some aspects of the evaluation process each year without central administration's urging (cf. O'Reilly, 1981; Zucker, 1981).³

Clout is needed, however, in those cases where there is resistance to evaluation activities and evaluation use, i.e., where even setting up and conducting the evaluation is likely to fail, to say nothing of the use of its results. The one obvious instance of this in our data was when it became necessary to call upon a superintendent numerous times to move an evaluation forward. This individual was charged with making a decision as to the fate of a controversial project. She was called into the evaluation process at two key points: first to make it clear to the teacher-coordinators that they must participate in the evaluation process; and then to require to all secondary English and math teachers in the system to complete a questionnaire. She also called and attended a meeting at which certain teachers filled in the most important evaluation questionnaire. Her presence clearly made a difference in the seriousness and effectiveness of that session. Even with her support, the evaluation encountered numerous difficulties and still today, the final report has not been disseminated. Clout may force participation, but cannot solve all the problems associated with a given evaluation.

Summary

This section of our report has had three purposes: first, to demonstrate that the suggestions for improving use contained in the evaluation literature are not panaceas and that in some cases the "failure of educa-

³Our thanks to Nick Smith for clarifying this point.

tional evaluation" in a local setting is a foregone conclusion; second, to suggest that blaming either users or evaluators in general for this is a counterproductive activity; and third, to extend the list of factors affecting evaluation use to include the notion of "espoused theories" vs. "theories-in-action," a revised personal factor, and the clout factor. These points are grounded in the LEA data collected this year, and, although our discussion of examples has necessarily been limited, additional supportive data are available and will be presented in a future publication. The next section of the report uses our grounded data to generate a conceptual framework for discussing evaluation use in local settings.

Section 4: A Conceptual Framework for Evaluation Use

Anyone who has ever worked with naturalistic data will readily admit that the process through which these data yield unifying concepts is a mysterious one. In our case, frequent staff meetings, readings of interview notes and available documents, and dialogues during the process of collecting new data led us to the occasional insights that ultimately provided the conceptual framework discussed below. While we acknowledge the danger of creating jargonish labels in any research project, we hope that the terms we use are helpful. We have used new terms only in those places where such labelling provides a lexical handle for previously unnamed concepts; those terms that are unneeded will presumably be eliminated in the course of future discussion and analysis.

The notion of grounded theory described in Glaser and Strauss (1967) provides the methodological approach for generating the concepts described in the following section. In this view, "grounded theory" consists of those

ideas that emerge directly from successive, intensive examinations and reconceptualizations of available data. It is "theory" because it seeks to conceptually explain observable experiences; it is "grounded" in that directly observed or experienced real world events, unscreened by a questionnaire filter, constitute the data base. In this initial analysis of our data, we have generated a conceptual framework with three dimensions: types of local use; the process of use; and the domain of the use process.

Types of Use

Three ideas from an earlier review of literature (King, Thompson, & Pechman, 1982) will frame the current discussion of use types:

1. Use vs. impact - In reviewing the literature, King, Thompson, and Pechman follow Smith (1980) in distinguishing between use and impact. The use of an evaluation is an intended act by an individual to achieve a given end, whereas an evaluation's impact occurs independently, without the planned intervention of a user.
2. Non-use as a form of use - The second framing idea from the earlier review is the notion that non-use can constitute a viable and appropriate form of use in the bureaucratic contexts characteristic of local settings.
3. Earlier definitions of the modes of use - Previous literature (see Knorr, 1977; Weiss, 1977; Pelz, 1978; Leviton & Hughes, 1979) has provided the following definitions: instrumental use, where an action corresponds directly to a given result; conceptual use, where over a period of time evaluation results influence a user's thinking about a problem; and symbolic or persuasive use, where users apply the evaluation process or its results for personal ends (e.g., to garner political support or to discredit a policy).

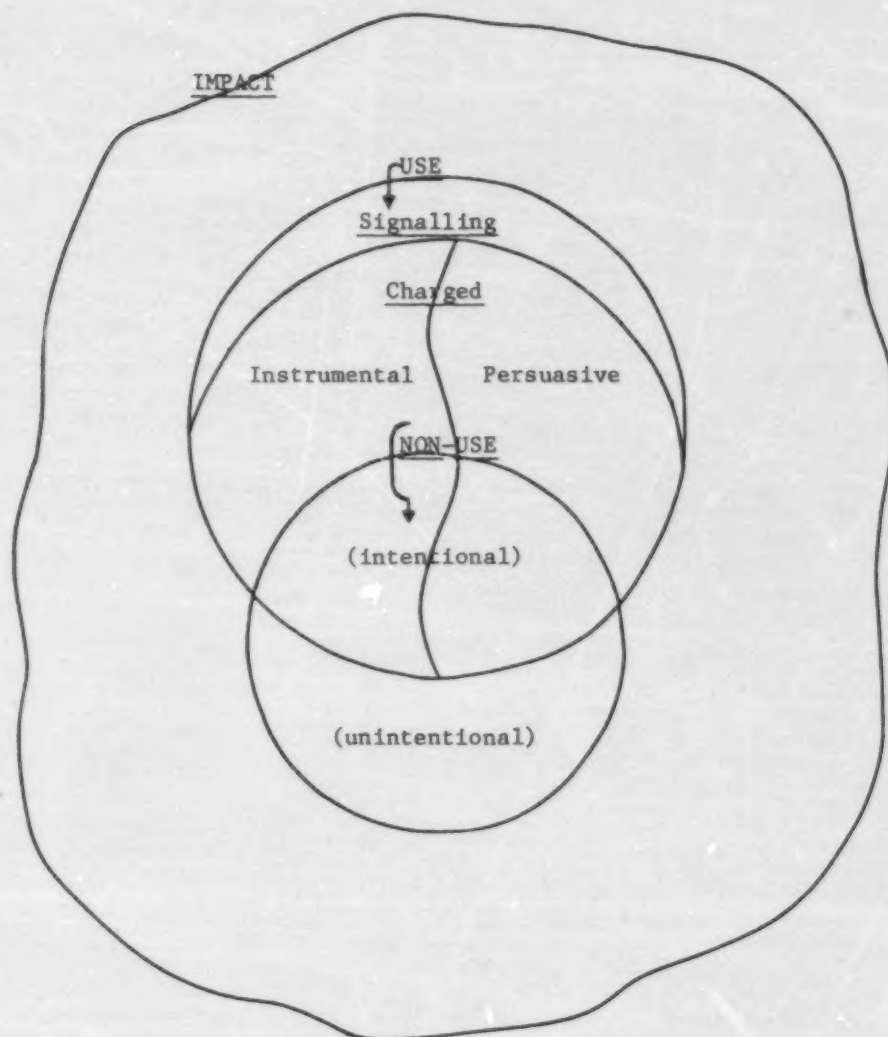
The analysis of our data has led us to re-conceptualize the types of use as presented in Figure 2. The basic distinction is between the unintentional outcomes of the evaluation process or its products--impact--and the intentional outcomes--what we are calling either "signalling" or "charged" use. Because our concern was with users' intentions, we will not discuss impact any more than to say that the effects of evaluations on the bureaucratic functioning of local school systems are many, varied, and worthy of additional study (see Bank & Williams, 1981).

Consider, then, the circle of intentional activity given in Figure 2. In examining the work of the R & E department, we immediately became aware that program evaluation as described in textbooks played only a small part in the unit's daily activities. More central to the system's purpose was the department's active involvement in collecting, processing, and distributing basic status information about the school district (e.g., state mandated statistical reports, student enrollment data and records, and state and federal evaluation reports). Until recently, this information and the activities that produced and maintained it formed the major responsibilities consuming most of the R & E staff's time.⁴ An awareness of the necessity of this information to the ongoing, smooth functioning of the system as a whole led us, in thinking about the use of evaluation information, to distinguish between information for use outside the system and information with a potential use inside the system.

The first type of use we are calling "signalling," after Zucker (1981). This refers to the use of evaluation information and activities as signals from the local school district to funding and legislative support agencies

⁴ As this report went to press, the functions of the R & E unit were substantially changed, in part as a result of this study, to place greater emphasis on direct and more traditionally conceived evaluation functions.

Figure 2. Types of Use: Signalling and Charged Use



that all is well. The information produced constitutes evidence signalling the successful fulfillment of legal obligations; it meets accountability expectations and requirements. Clearly, the use of evaluation information as an accountability signal is essential to the institutional survival of the school district, but it is usually not necessary to sustain the day-to-day routine of teachers and students at the school building level.

The necessary conditions for signalling, any one of which is not sufficient, are as follows:

1. The information consists of a tangible, deliverable product, usually a report;
2. It is created and delivered in a routinized and clearly defined process;
3. The use of the information takes place outside of the local system (and may well be unknown within the system); and
4. The forwarding of the information is required in the cyclic functioning of the school bureaucracy.

To label this function "use" at the local level is incorrect in that local people actually do very little with the information beyond packaging it neatly, according to the prescribed reporting formula, and forwarding it to the appropriate officials outside the system. The R & E staff serve as information middle-men whose job is complete once the information arrives properly formatted in someone's in-basket in the state capital or in Washington, D.C. While the way information is used by external groups may potentially be important, how the data are used is less critical to the R & E Department and to the school system than is the timely and accurate generation and reporting of the required information. What matters locally is that the reports are completed on time and according to specifications so

that the evaluation component of the funding proposal can be marked finished. The content of the report is often beside the point too because, as each of the evaluators we interviewed noted with somewhat cynical resignation, the federal funding and evaluation cycles are not synchronized. A good program's refunding or the termination of a weak program has little to do with the findings of final evaluation reports since they are usually submitted to funding agencies several months after refunding decisions have been made. The dilemma of the signalling function for the evaluation unit was summed up as follows by one of the R & E staff evaluators:

To some extent our role is like the factory worker who is responsible for inserting bolts in the engine. He is condemned if a bolt is missing from any engine. No matter how well he does his job no one notices him except when his job goes undone. The fewer engines that fall out of line, the more he is ignored. The same idea applies to R & E: if we do not stick the bolt in the right place, we are doomed; if we do, we are ignored.

From the local perspective, then, the use of evaluation information by state and federal funding sources seems binary in nature: if the prescribed report exists in the appropriate form, the program expenditure has been justified; if no report exists or if the correct reporting procedure is not followed, the use of the money will certainly be questioned. There may be little internal attention to the substance of these reports, and their use as data for program analysis or change is unusual. The signalling function of nearly all formal evaluations conducted in the system deserves special comment because, to the extent that evaluation reports are merely routine bureaucratic statements plugged back into the external funding system, their potential for effecting change is severely limited.

Figure 2 labels a second type of intentional use "charged." This notion developed from an examination of those cases where evaluation information was actively used by local managers. Like the ions in a chemical equation, evaluation information that takes on a charge--whether positive or negative--has the potential to cause a reaction in the system. Unlike the routine use of information for signalling, charged use carries with it the potential for disruption or change. Whereas signalling is institutionally oriented, charged use is person-specific; it provides an individual or group of individuals data upon which to base actions, and, to extend the metaphor, such use may occasionally cause sparks to fly. The charge may come from several sources: from the nature of the information itself, from an evaluator's presentation of the information, or from an individual user's concerns.

Unlike signalling use, the process that creates the information for charged use is neither clearly defined nor predetermined; furthermore, the information's form is unpredictable and may well be intangible. In our current thinking, charged use has only one condition, which is both necessary and sufficient: there must be evidence of the intentional and "serious consideration" (Leviton & Hughes, 1979) of evaluation information by an individual with the potential to act on it. Whether these individuals are called "idea champions" (Bank & Williams, 1981) or "illuminators" (Cronbach et al., 1980), they play a critical role in charged use, either adding the energy to charge existing information or to generate the needed information anew. As will be discussed later, the information so charged may or may not lead to an observable action. It is the attention paid that is the critical element.

Most of the data generated by the R & E Department we studied has the potential to become charged information. Two examples should make this point. The testing staff, for example, are frequently on the phone with parents, teachers, or administrators to answer questions about test procedures or results. Such data are "charged" because individual callers have specific needs, care about the results, and will see to it that the required follow-up occurs. A second, more specific example occurred when an administrator of a middle school curriculum program added a charge to the annual reporting of test results. As she put it,

I was tired of seeing high piles of printouts (of test scores) sitting unused in the resource room, so I decided this (a form for each student comparing old and new test scores) would be a way to get teachers to look at scores, to think about what might have happened to them.

Following consultation and in-service with the R & E Department's director, this administrator "charged" the routine test scores by focusing teachers' attention on individual students' needs as demonstrated by the test results and then used the results clinically.

While the primary division of the use circle in Figure 2 is between signalling and charged use, charged use is itself divided into two types: instrumental use and persuasive use. As was stated above, instrumental use occurs when an observable action can be directly linked to a specific piece of evaluation information. Our data suggest that this is a necessary, but not sufficient condition. A second such condition is that the user must consider the information prior to making his or her decision, i.e., the decision is not predetermined and the data are either collected or located explicitly for decision-making purposes. Instrumental use is a direct reaction to the evaluation information that emerges and for that reason is charged use; it is caused and has an effect. Alkin, Daillak, & White (1979) note the relative rarity of such use in local settings, although we did find

a few instances of instrumental use in our study. For example, when a superintendent this spring recommended eliminating a locally funded curriculum improvement program, she instrumentally used preliminary data that had been given to her at an informal meeting to argue that the program had not been effective. Her decision was made and implemented well before the final evaluation report could have been written. Other instances of instrumental use will be discussed below.

Because examples of instrumental use in our data were few, other evidence suggested the need for a second category of charged use called persuasive use, whereby a user selects information and puts it to use for his or her own ends. A single condition is both necessary and sufficient for persuasive use: the evaluation information is used (or not used) for what is believed to be personal gain. In contrast to instrumental use, the persuasive action taken is not necessarily suggested by the information itself; the information is, instead, sought or considered with a specific and personal end in sight, then used accordingly. This is not to say, however, that such use is necessarily malicious or conniving. It is merely part of the routine functioning of any bureaucracy. Consider, for example, the subject area supervisor who was fearful of cut-backs in his staff. During the year, he actively sought and broadcasted examples of the success of his programs, attempting to persuade others of their value and the need for continued financial support.

The wavy line dividing instrumental from persuasive use in Figure 2 is to note that the distinction may not always be as distinct as definitions require. If, for example, a decision-maker has a hunch about the questionable value of an unpopular program, she/he may set up an evaluation to collect damning data and then use the information to act accordingly. Is this, then, instrumental use because the action follows the direction implicit in the information, or persuasive use because the data were collected to

eliminate a program that had little support? Our experience and the evaluation literature (e.g., Patton, 1978) suggest that the results of evaluations should not surprise potential users in most cases, and the line between instrumental and persuasive use may at times be fine indeed and, in any case, certainly difficult to document.

Readers will note that the term symbolic use has been omitted from Figure 2. This is because in our observations what has been called symbolic use is clearly a subset of persuasive use. It is the user's effort to persuade others that the scientific spirit has been evoked, i.e., that an "evaluation" has been conducted, regardless of the content or quality of the process or products. For example, the previous head of one of the district's curriculum development programs repeatedly called for evaluations of the program. It became evident after three years, however, that the requests were symbolic because he effectively thwarted every attempt by the R & E unit to conduct a meaningful evaluation, while simultaneously explaining to others how much he believed in the process. Although symbolic use is thus a type of persuasive use, our feeling is that it may be less common in practice than we initially thought; we found few instances of such use in the system we studied and hence have eliminated it from our conceptualization as a separate use category.

It was noted earlier that non-use in many cases represents a viable form of use (Raizen & Rossi, 1981; King, Thompson, & Pechman, 1982), and Figure 2 therefore presents non-use with its own circle. Because non-use can be both intentional and unintentional, the circle is included both in use and in impact.

To count as an instance of instrumental non-use, users must intentionally consider evaluation information, but be unable or unwilling to act on the implicit recommendations. An example of this occurred in the federal project where the evaluation report stressed the urgent need of clerical services or at least a typewriter. The head of the project, the major program level decision-maker, would happily have used this evaluation data, but she had neither the money nor the power to obtain the resources and hence could not act. Her superior, however, could have used the information to take action, but chose not to for reasons of his own. The "hands-tied" nature of instrumental non-use may make it difficult to observe in bureaucratic settings where people are reluctant to reveal their weaknesses.

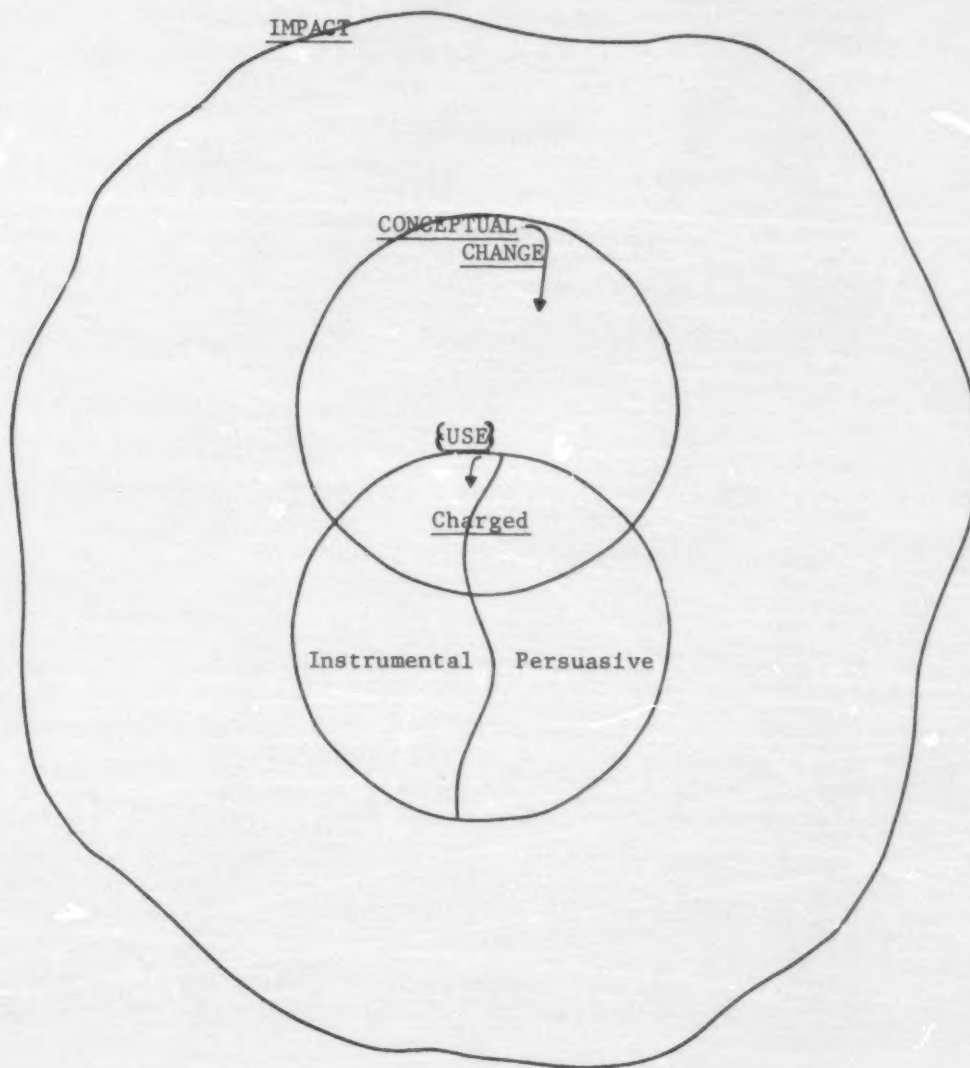
To count as an instance of persuasive non-use, the second type of charged non-use, users must intentionally decide not to apply information to decisions because they feel it is in their personal or program's best interest not to do so. So, for example, when the head of a federal program supported a thorough evaluation of his project and was then displeased by the results, his refusal to accept the figures as accurate was evidence of persuasive non-use. His reasons for not following certain direct recommendations contained in the final evaluation report were complicated and perhaps inappropriate, but they were part of the highly political context in which the program operates. A second example of persuasive non-use occurred when a building principal rejected positive information about a federal program in her building. Whereas she could have taken this as good news and used it as evidence of her school's success, she chose not to, instead using her negative personal feelings about the program and the program personnel to persuade her superiors of its poor quality.

In our conceptualization, unintentional non-use, a relatively common occurrence in local school settings, falls outside the use circle because it lacks the essential, intentional element that characterizes use.

One example is the nonuse of an evaluation of the achievement effects of small classes that was mandated by a superintendent whose functions changed during the year. The evaluation was designed and executed as the textbooks would recommend and then forwarded to the top administrators in the district in a timely manner. Because the superintendent who mandated the evaluation had changed his role in the school district, however, the evaluation and the project lost their "idea champion," and the data were never used. Others who were made aware of the evaluation showed no interest in considering either the report or the project's potential. A second example of unintentional non-use occurred when a superintendent received copies of evaluation reports and did nothing in response except perhaps file them--or misplace them on his desk--the information had in no sense been used. These instances are examples of evaluation impact, albeit rather negative ones, because evaluation information was received but set aside when it could have been used.

The remaining use concept from earlier research is presented, for simplicity's sake, in Figure 3. As typically defined, conceptual use lacks the intentional element of use; conceptual "use" is seen not as a conscious act on someone's part, but rather as the gradual change in someone's thinking over a period of time. For this reason, we call this activity conceptual change, rather than use. Of course, it is possible, as Chelimsky (1978) and Roecks (1982) note, to conduct a series of studies that may consciously lead to altered thinking about a given problem, and hence the conceptual change circle in part overlaps the use circle. In our year's data, not surprisingly, we documented no cases of conceptual change, suggesting the need for long-term, longitudinal study of this notion.

Figure 3. Types of Use: Conceptual Change



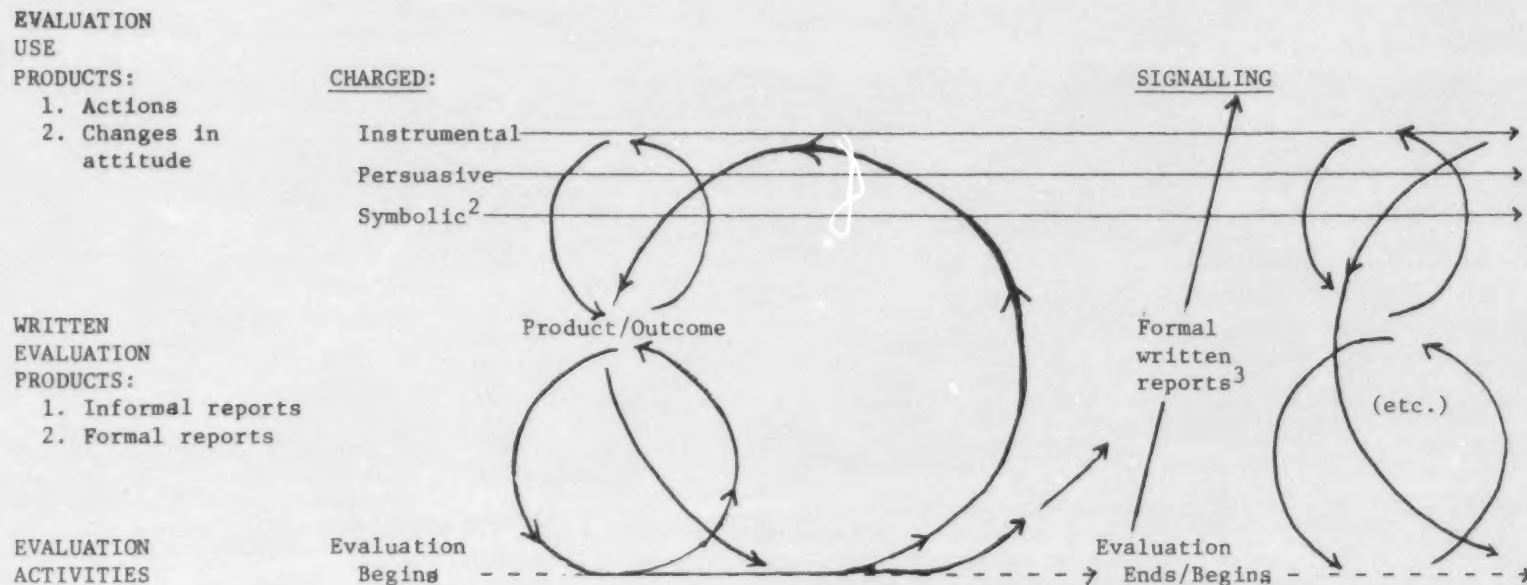
The Process of Use.

Based on the data collected this year and a reworking of earlier conceptual definitions, the use types presented in the previous section suggest the several ways in which local decision-makers use evaluation results. These types do not, though, capture the ongoing nature of the use process as we now see it, a process represented in Figure 4.

It is undeniable that this diagram greatly oversimplifies what it depicts. As we observed, the process of evaluation use in a school system couples multiple audiences and stake-holders in an intricate series of dialogues that begin even before the evaluation is undertaken and that usually continue long after the final report is written. Our notes document that people involved in the process may feel threatened, defensive, or upset; they may relish the opportunity to toot their own horn or to dent someone else's; they can welcome a brief slice of what they feel is objective truth or dismiss it as a compiling of hopeless inaccuracies. Because personalities are involved, the dynamics of the process change constantly, with earlier activities influencing the conduct and outcome of later events. However, like motherhood and apple pie, program evaluation is something people in school settings sense they ought to believe in. But what they mean when they smile and say they support evaluation varies from the top manager who seeks only good news about programs, to the supervisor annoyed at the lack of scientific rigor in district evaluations, to the principal eager for information that would help remove an unwanted federal project from her school. Tracing, capturing, and describing the process of use is akin to the classic dilemma of pinning a wave to the shore.

Nevertheless, Figure 4 diagrams our current sense of the use process. As can be seen, two components underlie the ongoing process: the evaluation activities themselves, i.e., the interactions that begin and end at some

Figure 4. The Process of Evaluation Use¹



¹Conceptual change is omitted from this figure only because it requires the input of information from a variety of sources over a longer period of time than is marked here.

²As is indicated here, symbolic use comes directly from the evaluation process and can even begin before the process itself. It requires no data-based information exchange whatsoever, only the existence of the process.

³This could be a mid-year or an end-of-year report.

point; and the obvious, tangible products generated by these activities, i.e., the evaluation's informal and formal written documents. The less tangible "products"--users' actions and attitudinal changes--do not underlie the use process; they are that process, representing charged use, signaling, or, over a period of time, conceptual change. Use is, after all, the ultimate evaluation product.

Figure 4 suggests that use originates in one of two ways. First, the repeated formal and informal interactions between evaluator and potential user during evaluation activities can in and of themselves lead directly to use. Brief examples from the LEA make this clear. When an evaluator merely asked the director of a federal project why certain assistant teachers had attended an in-service workshop, the director immediately realized that their attendance was inappropriate, and they were not included in the next such session. Or, when another federal project director was told informally that newly purchased texts and supplies were unused and kept locked in a closet, he acted to make the materials available to children.

A second way in which use begins is in response to a written evaluation product. What is labelled a product can range from a handwritten memo documenting a casual observation to a formally typed and bound end-of-year report sent to the State Department of Education. The critical point is that once information is put down on paper, it takes on a life of its own within and outside the school bureaucracy; it becomes important and potentially dangerous. Consider an example from a small federal project operating in a single school.

The project director, who was also responsible for teaching in the project, had been left on her own, her project supervisor making it clear that the project was hers and hers alone. Her repeated attempts to get essential clerical help or even a typewriter to type curriculum materials

were to no avail. Because an evaluation was mandated, with R & E's help she hired an external evaluator, who collected data documenting her difficulties in implementing the project without clerical assistance. The mid-year report emphasized this, calling on the appropriate administrators to arrange for the needed help. Several months later, the teacher received a memorandum from the cited "official," cc'd to seven people (up to the level of deputy superintendent) and containing the following comments:

You stated some time ago that two (2) months were needed... in order to complete the project. The fact that clerical help was curtailed due to budgetary cutbacks or unforeseen circumstances does not in any way relate to this understanding...

If the project is not completed it will be due to improper, incomplete, poor planning; not providing for unforeseen circumstances or not foreseeing all circumstances which could come up...

I believe it is my duty to inform you of my opinion concerning completion of your project since I and the Project Supervisor are mentioned in the Evaluator's report (emphasis added).

The administrator's reason for writing the memo is evident: in the unlikely event that anyone should read the report, he is seeing to it that he will not be blamed for the project's failure, despite the fact that he had been unwilling to provide clerical help (beyond saying that legible handwritten materials turned in on time would be acceptable). This case and others like it demonstrate the power of the written word in a school bureaucracy.

In our data, examples of use flowing from written evaluation products are relatively common. One superintendent with a mandate from her superior to make substantive changes in her division requested R & E's help with a needs assessment, noting that the principals "will believe printouts" where they might not believe her. She used the data collected to help implement the service changes she wanted. As mentioned previously, a second superin-

tendent recommended the elimination of a staff development and curriculum improvement program based on information in an interim informal progress report. A federal project director used documented evaluation results to sell his terminating project to the system, and so on. Evaluation products were used both instrumentally and persuasively in the system we studied.

Figure 4 also points to the relatively passive role of the signalling function in local evaluation use. Undeniably critical though that role is, the signalling happens only once or twice per evaluation, relies only on formal written reports--certainly minimal pieces of the entire set of evaluation activities--and rarely affects the program directly. This figure clarifies the limited perspective of the earliest studies of evaluation use and helps explain why users repeatedly report that they are using evaluation data; there are many ways they can do so.

The literature has recognized that the process of evaluation use is highly complex and oftentimes difficult to trace. However, our data and Figure 4 suggest three important points. First, the use process can begin before the evaluation process and certainly can continue well after the evaluation process ends. Second, regardless of the type of use, the use process is ongoing and can repeatedly pull information from evaluation activities directly, as well as from the products they generate. And third, the signalling function represents only a small subset of the possible uses to which evaluation information can be put.

The Domain of Charged Use.

Throughout our study, we discovered that many users, having little background in measurement or evaluation, felt encouraged to use what appear to be "hard data" or research-based information, but that such data are often less objective than they may seem. Nevertheless, to an observer such inappropriate use may appear good simply because data were used in some way

to affect decisions. Others have noted the danger of assuming that all use is good and all non-use bad (e.g., Caplan, 1980), and the question for us became how to distinguish between appropriate and inappropriate use, our goal as evaluators, after all, being to maximize the first and eliminate the second.

The development of Figure 4 helped us in thinking about this question because once we realized that evaluation activities and written products together triggered the use process, we could then connect specific activities, products, and instances of use (i.e., specific behaviors or attitude changes) in "equations"⁵ of the form,

$(\neg) A \rightarrow (\neg) (B) \rightarrow (\neg) C$, where
 A represents appropriate evaluation activities;
 B represents an appropriate evaluation product;
 C represents appropriate use; and
 The symbol $\neg$ negates or marks as inappropriate the letter it precedes.

As in linguistic notation, the parentheses denote an optional element. So, for example, the "ideal" case, $A \rightarrow B \rightarrow C$ may occur relatively infrequently because the evaluation product (B) may not be necessary. As shown earlier, use can proceed directly from the evaluation process (i.e., where $A \rightarrow C$) and will often occur prior to the completion of the product. A contrasting example, $A \rightarrow B \rightarrow \neg C$, represents an inappropriate instance of use following appropriate evaluation activities and products (e.g., where a user distorts data for personal gain), while $A \rightarrow \neg B \rightarrow \neg C$ represents an inappropriate instance of use stemming from appropriate activities, but a poor product (e.g., where a user seriously applies reported misinterpreta-

⁵ The term equation smacks too strongly of empiricism and quantification for our liking, but we know of no better label for such linkings of elements.

tions of test results to make curricular decisions). More generally, we feel that use (C) will be appropriate only in instances where both the activities (A) and the products (B) are also appropriate.

But how exactly does one decide the appropriateness of any of these elements? Because such labelling requires us to make professional value judgments, one's perspective of the entire evaluation process (e.g., as user or evaluator) clearly influences the final decision to label something as yes, appropriate or no, inappropriate. Nevertheless, such judging and labelling occur frequently in education, and the existence of the Joint Committee's Standards (1981) provides an accepted framework for rating evaluation activities and products. Difficult though it may be, the value of this equation approach for analyzing instances of use is that it enables us--indeed, it forces us--to link the use process with its associated evaluation activities and products. In this way evaluators and researchers alike will have a way both to analyze their wins and losses and to work on ways to increase their success rate.

Figure 5 summarizes the domain of potential instances of use, adding the variable of non-use, that, as was mentioned earlier, must be considered in any discussion of use. In the figure, use of any sort is represented by C_1 , non-use by C_2 , and mis-use by either $\neg C_1$ (inappropriate use) or $\neg C_2$ (inappropriate non-use⁶). As the figure shows, there are eight possible cases of use/non-use. The first four occur where both the evaluation and products (if they exist) are "good."

Case 1, appropriate use, in which an appropriate product is used wisely;

⁶ Again, the chart looks more complicated and fixed than it is meant to.

Figure 5. The Domain of Charged Use: Use (C_1), Non-use (C_2), and Mis-use ($\neg C_1$ or $\neg C_2$)

a. Appropriate activity/product: $A \rightarrow (B) \rightarrow ?$

	Appropriate	Inappropriate
Use	Case 1 $C_{1.0}$ (Appropriate use of appropriate product)	Case 2 $\neg C_{1.1}$ (Manipulation or other inapprop. use of appropriate product)
Non-use	Case 3 $C_{2.0}$ (Appropriate product cannot be used for reasons beyond user's control)	Case 4 $\neg C_{2.1}$ (Appropriate product consciously ignored) $\neg C_{2.2}$ (User indifferent to appropriate product) $\neg C_{2.3}$ (User unaware of appropriate product)

b. Appropriate activity/inappropriate product or inappropriate activity/product:

$$\left\{ \begin{array}{l} A \rightarrow \neg B \\ \neg A \rightarrow (\neg B) \end{array} \right\} \rightarrow ?$$

	Appropriate	Inappropriate
Use	Case 7 (Throw out the evaluation; fire the evaluator)*	Case 5 $\neg C_{1.2}$ (Inapprop. use of inapprop. activity or product)
Non-Use	Case 6 $C_{2.5}$ (Appropriate <u>non</u>-use of inapprop. activity or product)	Case 8 (Using the inapprop. product and <u>not</u> firing the evaluator)*

*Within a program, inappropriate products cannot be used appropriately or not used inappropriately; however, appropriate action in Case 7 would be to throw out the evaluation and to fire the evaluator. In Case 8, inappropriate non-use would consist of not throwing out the evaluation and firing the evaluator.

Case 2, inappropriate use, in which an appropriate product is somehow misused;

Case 3, appropriate non-use, in which, for reasons beyond a user's control, an appropriate product cannot be used; and

Case 4, inappropriate non-use, in which an appropriate product is either purposely or carelessly ignored.

The remaining four cases result either when appropriate activities lead to inappropriate products or when both the activities and products are inappropriate:

Case 5, inappropriate use in which appropriate activities or products are unwisely used;

Case 6, appropriate non-use, in which inappropriate activities or products are wisely not used;

Case 7, appropriate use, in which inappropriate evaluation activities or products lead either to firing the evaluator or to a new evaluation;

Case 8, inappropriate non-use, in which poor evaluation activities or products were ignored, i.e., where no action was taken to fire the evaluator or to redo the evaluation.⁷

Taken together, these cases suggest the following points:

1. Non-use can be appropriate in two instances (cases 3 and 6), i.e., when the results don't deserve to be used or when the evaluation's organizational or political context unavoidably prevents a user from acting on an appropriate product.

⁷ Our thanks to Nick Smith for suggesting the examples in cases 7 and 8.

2. One type of mis-use, inappropriate use, can occur in one of two ways (cases 2 and 5), i.e., either a user manipulates or otherwise mis-applies appropriate results or s/he inappropriately makes use of an inappropriate product.

3. A second type of mis-use, inappropriate non-use, can occur in several ways (cases 4 and 8), i.e., the results are ignored because a potential user chooses to ignore them maliciously, is simply indifferent to them, or doesn't even know about them; or poor evaluation activities and products are inappropriately allowed to stand unchallenged and unquestioned.

4. There are two likely causes of misuse (cases 2, 4, 5, and 8), i.e., the inappropriate use (or non-use) of evaluation results: either a certain level of incompetence on the user's part (using inappropriate data, being unaware of appropriate data) or his or her strategic sensitivity to organizational politics (manipulating or ignoring appropriate or inappropriate data for political reasons).⁸

To understand these ideas without having actual examples would be difficult, so Table 1 summarizes the ideas in a different form, as well as giving examples of each use type from our study. The relative nature of these labels when applied to real instances must be noted; what to one person is mis-use may be, from another person's perspective, perfectly appropriate use.⁹ But, as noted earlier, the value of this framework lies in its linking of evaluation activities, products, and use. Rather than

⁸ Our thanks to David Ransen for first mentioning the fourth point and to Marv Alkin for clarifying it.

⁹ A good example of this would be the memorandum quoted on page 50; the person who wrote it no doubt felt he was responding appropriately to the evaluation report he had read.

Table 1. The Eight Possible Cases of Charged Evaluation Use

Case	Evaluation Activities: Appr. or Inapp.?	Written Products: Appr. or Inapp.?	Evaluation Use: Appr. or Inapp.?	Type of Use and Use Equation	Examples from Study
1	Appr.	(Appr.)	Appr.	Appropriate use $A \rightarrow (B) \rightarrow C_{1.0}$	Federal project director receives informal feedback from evaluator and alters project accordingly Superintendent cuts a staff development and curriculum improvement program after receiving a memorandum of evaluation data
2	Appr.	(Appr.)	Inapp.	Inappropriate use $A \rightarrow (B) \rightarrow C_{1.1}$	Instructional supervisor writes a lengthy memo denying responsibility for the implementation problems of a federal project under his jurisdiction
3	Appr.	(Appr.)	Appr.	Appropriate non-use $A \rightarrow (B) \rightarrow C_{2.0}$	Because a program as a whole has been cut, a superintendent is unable to implement recommendations from its evaluation
4	Appr.	(Appr.)	Inapp.	Inappropriate non-use $A \rightarrow (B) \rightarrow C_{2.1-2.3*}$	A successful federal project's funding period ends, and the program is allowed to die, despite two highly favorable evaluations A superintendent simply ignores the results of a carefully conducted needs assessment he requested

* It is impossible to distinguish among $C_{2.1-2.3}$ in actual instances because to do so would be to attribute motivations we simply cannot know of to potential users.

Table 1. (Continued)

Case	Evaluation Activities: Appr. or Inapp.?	Written Products: Appr. or Inapp.?	Evaluation Use: Appr. or Inapp.?	Type of Use and Use Equation	Examples from Study
5	a. Appr. b. Inapp.	a. Inapp. b. (Inapp.)	Inapp.	Inappropriate use $\left\{ \begin{array}{l} A \rightarrow \neg B \\ \neg A \rightarrow (\neg B) \end{array} \right\} \rightarrow \neg C_{1.2}$	Project staff from one federal project apply the results of a poorly designed, self-conducted study in planning for the following year
6	a. Appr. b. Inapp.	a. Inapp. b. (Inapp.)	Appr.	Appropriate non-use $\left\{ \begin{array}{l} A \rightarrow \neg B \\ \neg A \rightarrow (\neg B) \end{array} \right\} \rightarrow C_{2.5}$	A new superintendent ignores a superficial and hastily prepared report about a major curriculum improvement program A poorly written evaluation report with no recommendations for change is ignored by a federal project director
7	a. Appr. b. Inapp.	a. Inapp. b. (Inapp.)	Appr.	Appropriate use $\left\{ \begin{array}{l} A \rightarrow \neg B \\ \neg A \rightarrow (\neg B) \end{array} \right\} \rightarrow **$	None (Theoretical example: fire the evaluator; re-do the evaluation)
8	a. Appr. b. Inapp.	a. Inapp. b. (Inapp.)	Inapp.	Inappropriate non-use $\left\{ \begin{array}{l} A \rightarrow \neg B \\ \neg A \rightarrow (\neg B) \end{array} \right\} \rightarrow **$	None (Theoretical example: don't fire the evaluator; don't re-do the evaluation)

** As explained in the note on Figure 5, inappropriate products cannot be used appropriately or not used inappropriately within a program.

being separate from the ongoing events of evaluations, use and non-use can now be seen as their natural extension.

Although this initial report has not presented many of the examples we have available, the three sections of this conceptual framework for evaluation use have nevertheless relied heavily on the data collected this year in a large city LEA to give the range of evidence we gathered in a succinct format. The types of use section modified earlier ideas based on the kind of use we actually observed; the process of use section connected ongoing evaluation activities and written products to the evaluation products we call use--users' actions and attitude changes--because we watched this happen; and the domain of charged use section was our first attempt at categorizing the variety of instances of use, non-use, and mis-use we collected. This framework, integrated with the factors contributing to evaluation use discussed in Section 3, represents the major concepts generated in this year's study of the process of school evaluation use. Later analyses will include more of our data and may somewhat alter this conceptual framework, but we have presented here material we hope will stimulate discussion in the field. The final section of this report will discuss the implications these findings have for increasing appropriate evaluation use in schools.

Section 5: Summary and Implications

Summary

In the introduction to this report, we listed two questions as central to our study: (1) In what ways do local decision-makers use evaluation information generated by the research and evaluation unit of a large public school system? and (2) What concepts can describe the use of such information within an LEA? It should now be obvious that the answers to these questions have to date been only partially answered. Empirical research on the local use of educational evaluations has provided some helpful information in recent years. Evaluators should no longer expect users to automatically plug the results of evaluations into their routine decision-making, nor should they expect to find predictable behavior that could necessarily be labelled "rational" outside of a given organizational context. Data gathered for this project have also suggested that evaluators should view improved evaluation reports and ongoing collaboration with users as partial remedies that may work in some instances, but not in others.

This is not to say, however, that responsibility must be placed on someone's shoulders for the field's pervasive sense of the non-use of results. While in specific instances individuals can easily be barriers to evaluation use, in general it is neither users as a group nor evaluators as a group who are at fault. Rather, it is the process of evaluation use itself, and the use of that process in local settings, that may lead to situations where non-use is the only possible outcome. By better understanding this process, evaluators may be able to determine in advance likely problem areas and strategies for resolving them.

This report has presented three additional or revised factors to the list of those affecting local use. First, in our discussions with users, we

frequently sensed a distinction between what they said they believed (their "espoused theory") and what they really believed, given their actions (their "theory-in-use"); this may be an important factor determining how people actually respond to evaluations, regardless of what they say. Second, in those instances where use occurred, we typically found a user who was self-confident and open to the information, regardless of its nature. The personal factor, as discussed in the literature, may need to expand to include this dimension. The third factor we observed affecting local use was the clout factor--the support of evaluation activities by key administrators in the system. While this may be important only in those cases where participants resist evaluation, its effect can, nevertheless, be powerful.

Our report has also described a grounded conceptual framework for discussing evaluation use as we define it. Use is marked by the conscious selection and application of information, and the framework has three component concepts:

1. The types of use category distinguishes first between "signalling," whereby information is sent out of the system to signal that required activities are proceeding as mandated, and "charged use," whereby a user actively takes information and does something with it. Charged use is further divided into instrumental use, an observable action that can be linked to a specific piece of information, and persuasive use, use of information by an individual for his or her own ends. Conceptual change, a third "use" type, typically lacks the conscious element characteristic of use and in most cases represents the indirect impact of a series of evaluations, rather than their use.

2. The second concept in our framework is the process of use, as diagrammed in Figure 4. Evaluation activities and written products underlie this process, which itself can have two products--users' actions and changes

in attitude. Regardless of the type of use, the ongoing and dynamic process can repeatedly pull information both from evaluation activities and their products. Use may begin symbolically before any evaluation activities start and may also continue after all have ended. In this context, the minimal role of the signalling function can be easily seen because although written information may be the longest lasting, most tangible product of evaluations, the elusive, nonquantifiable process of evaluation is typically more important for use.

3. The domain of charged use, the third element in the framework, presents eight possible cases of use, combining evaluation activities, written products (if they exist), and use, each judged as appropriate or inappropriate, in "equations" defining appropriate use and non-use, and inappropriate use and non-use (see Figure 5 and Table 1). Mis-use is defined as either inappropriate use or non-use. Coupling the process of use with value judgments as to appropriateness, these use equations enable us to determine those instances where use is "good" and those where it is not.

Advice to Evaluators

What do these results say to practicing evaluators? Most LEA evaluators may already be well aware of the implications of what is written here, but at the risk of stating the obvious, we would make the following suggestions. First, regardless of their setting, evaluators should do all that they can to conduct the highest quality evaluations possible. The fact that local conditions often conspire against the process makes this a difficult charge, but, if we give up and accept mediocrity, our purpose will be lost. Professionalism demands no less.

This is not to say, however, that the evaluation enterprise must be conducted blindly, impervious to the political and interpersonal factors

that daily affect its ongoing dynamics. Our second piece of advice, then, is for evaluators to become highly sensitive to the evaluation contexts in which they work. This includes the careful study of potential users, i.e., differentiating between people's "espoused" evaluation theories and their "theories-in-use" actions, analyzing decision-makers for the self-confidence and openness to information that would make them high probability users and then responding accordingly.

Being context-sensitive also involves assessing the politics of each situation. Evaluators need to recognize those situations where the likelihood of their failure is high. If there is obvious resistance to the evaluation and clout can be used, it should be positively applied; if clout is not forthcoming, then the evaluator must work creatively within the given constraints, taking every opportunity to demonstrate the limitations of an evaluation product under the circumstances. Granted, this is no easy task and runs the risk of undermining the evaluator's standing in the bureaucracy. Hopefully, over a period of time, administrators with the needed clout will begin to appreciate the demands of an effective evaluation process and provide appropriate support. In any case, evaluators who are skillful at assessing the personal and political climate in which they work should experience fewer surprises and be better able to prepare responses proactively rather than reactively.

The conceptual framework we presented is less directed to the practicing evaluator than to evaluation theorists. Evaluators in the field no doubt understand the process of evaluation, having lived it from the inside out; and the use equations may be of value only to label similar cases they have experienced. Two additional recommendations can be made, however. First, evaluators should recognize when their work is intended only for signalling use. This should create a set of expectations different from

instances where evaluations are designed to be used by someone in the system. Second, recognizing the nature of charged use, whether instrumental or persuasive, evaluators should work with decision-makers to increase use. While we know that this will not work in every case, it is possible that over time, conceptual change will occur and enable these users to take advantage of the evaluation process. This report has suggested that while there are not guarantees, the attempt in and of itself is worthwhile.

Thoughts on Future Research

The study discussed in this report documents a relatively rare research opportunity, a true participant-observer/observer-participant study. Based upon our experience, we are convinced of the appropriateness of using naturalistic methods to capture the ongoing process of evaluation use. Actually living and working with a beleaguered R & E unit helped us focus on the evaluation process as it occurred in a local setting. Instance chasing caused us to track examples of use from multiple perspectives and to focus on the contexts of specific evaluations. Repeated interviews with both R & E staff and local users provided detailed information about how certain individuals viewed their roles in creating or preventing use, and ongoing project staff discussions helped us clarify research concerns and locate common themes in our data.

Our first recommendation for additional research, then, is methodological, i.e., that other investigators apply naturalistic methods in a variety of evaluation settings in order to build substantial data bases for subsequent analysis. This suggestion is not original (see e.g., Alkin, 1979) and, although the time for quantitative studies of evaluation use is also upon us, the richness of naturalistic data for suggesting new and critical elements of the use process is undeniable. As researchers, we must

learn to keep our eyes and ears open to the subtleties of events occurring in evaluation contexts, and naturalistic methods teach us to do so.

A second recommendation, concerning the factors and concepts posited in this report, is for the testing and subsequent confirmation or refutation of these ideas. We need to ask questions like the following: 1) Are the types of use as we describe them accurate and complete? 2) Is our presentation of the process of use typical of other evaluation settings, in LEA's and elsewhere? 3) Do the equations delimiting the domain of the use process clarify or muddy efforts at increased understanding?

If the second recommendation would encourage continued theory-building about evaluation use, the third is more practical, seeking specific ways to optimize the appropriate use of the evaluation process in LEA's. The data discussed here suggest that evaluations in some cases will necessarily fail. If this is the case, what can local evaluators do to increase the likelihood of their success? Can evaluability assessments be useful in local school settings, and are they a realistic option? Can users' personalities somehow be "measured" to determine their openness to information? Can users be "forced" to use information? Can evaluators learn to distinguish counter-productive espoused theories and theories-in-use? On a different level, who, finally, should pay for evaluations? Who should control the process and outcomes? And how do these issues relate to the functioning and funding of the R & E unit within the school organization as a whole? What is an appropriate role and organizational position for the unit? The more we learn about evaluation use, the larger the number of unanswered questions grows. These do, however, provide a focus for further work.

A final and related research recommendation concerns the use, if any, of evaluation information at the classroom level. Kennedy, Apling, and Neumann (1980) and Daillak (1980) note that such use is rare, despite the

widely accepted notion that good evaluation should lead ultimately to better learning in classrooms. The need exists, therefore, to study the ongoing but informal evaluation process teachers and principals use in their day-to-day work. Comparing this to the information on the use of more formal evaluation results may provide valuable information about both processes (see Stecher, Alkin, & Flescher, 1981).

A Final Comment

It should be clear by now that the ideas presented in this report represent only an initial analysis of reams of data collected an LEA during an entire school year. Over the course of the next year, we will work with these data to clarify the ideas presented here and, perhaps, to add others. What we have learned thus far is that certain common sense hunches for improving evaluation use will not guarantee automatic success, and the concepts developed try to suggest reasons why this may be the case. Our final goal is to provide information that will lead to improved LEA evaluation use, and it is with this in mind that we return to work.

References

Alkin, M.C. Evaluation: Who needs it? Who cares? Studies in Educational Evaluation, 1975, 1, 201-212.

Alkin, M.C. Using naturalistic research for the study of evaluation utilization. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979. (ERIC Document Reproduction Service No. ED 170 378)

Alkin, M.C., Daillak, R.H., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills: Sage, 1979.

Alkin, M.C., Stecher, B.M., & Geiger, F.L. Title I evaluation: Utility and factors influencing use. Northridge, CA: Educational Evaluation Associates, 1982.

Argyris, C. & Schon, D.A. Theory in practice: Increasing professional effectiveness. San Francisco: Jossey-Bass, 1974.

Bank, A., & Williams, R.C. Evaluation in school districts: Organizational perspectives. CSE Monograph Series in Evaluation, No. 10. Los Angeles: Center for the Study of Evaluation, 1981.

Brown, R.D., & Braskamp, L.A. Summary: Common themes and a checklist. In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Caplan, N. The two communities: Theory and knowledge utilization. American Behavioral Scientist, 1979, 22, 459-469.

Caplan, N. What do we know about knowledge utilization? In L.A. Braskamp & R.D. Brown (Eds.), Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1980.

Daillak, R.H. A field study of evaluators at work. Los Angeles: Center for the Study of Evaluation, 1980.

Datta, L. O thou that bringest the tidings to lions: Reporting the findings of educational evaluations. Paper presented at the annual Johns Hopkins University National Symposium on Educational Research, Baltimore, 1979.

Denny, T. How might evaluation reports be? Paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Dickey, B. Utilization of evaluations of small scale educational projects. Educational Evaluation and Policy Analysis, 1980, 2, 65-77.

Glaser, B., & Strauss, A. The Discovery of grounded theory. Strategies for qualitative research. Chicago: Aldine Publishing Company, 1967.

Guba, E.G., Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series in Evaluation, No. 8. Los Angeles: Center for the Study of Evaluation, 1978.

Guba, E.G., & Lincoln, Y.S. Effective evaluation. San Francisco: Jossey-Bass, 1981.

Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill, 1981.

Kennedy, M.M. Working knowledge and other essays. Cambridge, MA: The Huron Institute, 1982.

Kennedy, M.M., Apling, R., & Neumann, W.F. The role of evaluation and test information in public schools. Cambridge, MA: The Huron Institute, 1980.

King, J.A., & Thompson, B., "A nationwide survey of administrators' perceptions of program evaluation," NASSP Bulletin, in press (a).

King, J.A., & Thompson, B., "Research on school use of program evaluation: A literature review and research agenda," Studies in Educational Evaluation, in press (b).

King, J.A., Thompson, B., & Pechman, E.M. Evaluation utilization: A bibliography. New Orleans: Orleans Parish Public Schools, 1981. (ERIC Document Reproduction Service No. ED 207 984) (a).

King, J.A., Thompson, B., & Pechman, E.M. Evaluation utilization: An annotated bibliography. New Orleans: Orleans Parish Public Schools, 1981. (ERIC Document Reproduction Service No. ED 204 363) (b).

King, J.A., Thompson, B., & Pechman, E.M. Improving evaluation use in local school settings. A revision of the Final Report of NIE Grant NIE-G-80-0082. New Orleans, LA: New Orleans Public Schools, 1982.

Knorr, K.D. Policymakers' use of social science knowledge: Symbolic or instrumental? In C.H. Weiss (Ed.), Using social research in public policymaking. Lexington, MA: Heath, 1977.

Lee, A.M., & Holley, F.M. Communicating evaluation information: Some practical tips that work. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

Leviton, L.C., & Hughes, E.F.X. Utilization of evaluation: A review and synthesis. Evanston, IL: Center for Health Services and Policy Research, Northwestern University, 1979; also published as Research on the utilization of evaluations: A review and synthesis, Evaluation Review, 1981, 5, 525-548.

O'Reilly, C., III. Evaluation information and decision-making in organizations: Some constraints on the utilization of evaluation research. In Bank, A. & Williams, R.C., (Eds.), Evaluation in school districts: Organizational perspectives. CSE Monograph Series in Evaluation, No. 10. Los Angeles: Center for the Study of Evaluation, 1981.

Patton, M.Q. Utilization-focused evaluation. Beverly Hills: Sage, 1978.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., Grench, B.D., & Blyth, D.A. In search of impact: An analysis of utilization of federal health evaluation research. In C.H. Weiss (Ed.), Using social research in public policymaking. Lexington, MA: Heath, 1977.

Pechman, E.M., & King, J.A. Using Information: A description of the functions of a department of research and evaluation. Unpublished manuscript, 1982.

Pelz, D.C. Some expanded perspectives on use of social science in public policy. In J.M. Yinger & S.J. Cutler (Eds.), Major social issues, a multidisciplinary view. New York: The Free Press, 1978.

Raizen, S.A., & Rossi, P.H. Program evaluation in education: When? how? to what ends?. Washington, D.C.: National Academy Press, 1981.

Roecks, A. Comments on a draft of "The process of evaluation use in local school settings," personal communication, August, 1982.

Rubin-Frankel, M. A strategy for improving local educational evaluation uses in a southern LEA. Unpublished manuscript, 1981.

Rutman, L. Planning useful evaluations. Beverly Hills: Sage, 1980.

Smith, N.L. The president's corner. Evaluation News, Summer, 1980, 24-25.

Smith, N.L. Comments on a draft of "The process of evaluation use in local school settings," personal communication, August, 1982.

Stake, R. Evaluation design, instrumentation, data collection and analysis of data. Urbana, IL: Center for Instructional Research and Curriculum Evaluation, 1976.

Stecher, B.M., Alkin, M.C., & Flesher, G. Patterns of information use in school level decision-making. Los Angeles: Center for the Study of Evaluation, 1981.

Thompson, B., & King, J.A. Evaluator types: Do differences matter? Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

Weiss, C.H. (Ed.). Using social research in public policy making. Lexington, MA: Lexington Books, 1977.

Weiss, C.H. Conceptual issues in measuring the utilization of research and evaluation. Paper presented at the annual meeting of the Evaluation Research Society, Minneapolis, 1979.

Wise, R.I. What we know about the decision-maker and decision settings. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

Zucker, L.G., Institutional structure and organizational processes: The role of evaluation units in schools. In Bank, A., & Williams, R.C. (Eds), Evaluation in school districts: Organizational perspectives. CSE Monograph Series in Evaluation, No. 10. Los Angeles: Center for the Study of Evaluation, 1981.